软件学报 ISSN 1000-9825, CODEN RUXUEW
2026,37(12):1−26 [doi: 10.13328/j.cnki.jos.007771] [CSTR: 32375.14.jos.007771]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

中文生成式引擎优化*

李 彤[1,2,6], 吴荣榜[2], 李清昊[3], 贾 贵[3], 高翔宇[4], 范 举[1,2], 徐 恪[5]

[1](数据工程与知识工程教育部重点实验室 (中国人民大学), 北京 100872)

[2](中国人民大学 信息学院, 北京 100872)

[3](湖南牛投邦科技有限公司, 湖南 长沙 410023)

[4](清华大学 网络科学与网络空间研究院, 北京 100084)

[5](清华大学 计算机科学与技术系, 北京 100084)

[6](互联网体系结构全国重点实验室 (清华大学), 北京 100084)

通信作者: 范举, E-mail: fanj@ruc.edu.cn

摘 要: 大语言模型 (large language model, LLM) 和生成式引擎正在改变用户获取网络信息的方式. 传统搜索引擎优化主要关注网页能否进入搜索结果并获得较高排名, 而生成式引擎优化则关注网页内容能否被大语言模型采用并呈现在回答中. 然而, 现有研究主要面向英文场景, 面向中文网页和国内生成式平台的系统性生成式引擎优化 (generative engine optimization, GEO) 研究仍然缺失. 为此, 构建面向中文场景的生成式引擎优化评测框架, 围绕中文数据集构建、生成式引擎优化策略设计和来源可见度评测开展研究, 并从策略、平台、领域和来源位置等方面分析中文生成式引擎优化的效果. 在此基础上, 进一步比较中文生成式引擎优化与英文生成式引擎优化的差异, 并总结面向中文场景的实践启示.

关键词: 大语言模型; 搜索引擎优化; 网页内容改写; 引用分析; 跨平台评测

中图法分类号: TP311

中文引用格式: 李彤, 吴荣榜, 李清昊, 贾贵, 高翔宇, 范举, 徐恪. 中文生成式引擎优化. 软件学报, 2026, 37(12): 1–26. http://www.jos.org.cn/1000-9825/7771.htm

英文引用格式: Li T, Wu RB, Li QH, Jia G, Gao XY, Fan J, Xu K. Chinese Generative Engine Optimization. Ruan Jian Xue Bao/Journal of Software, 2026, 37(12): 1–26 (in Chinese). http://www.jos.org.cn/1000-9825/7771.htm

Chinese Generative Engine Optimization

LI Tong[1,2,6], WU Rong-Bang[2], LI Qing-Hao[3], JIA Gui[3], GAO Xiang-Yu[4], FAN Ju[1,2], XU Ke[5]

[1](Key Laboratory of Data Engineering and Knowledge Engineering (Renmin University of China), Ministry of Education, Beijing 100872, China)

[2](School of Information, Renmin University of China, Beijing 100872, China)

[3](Hunan Niutoubang Technology Co. Ltd., Changsha 410023, China)

[4](Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing 100084, China)

[5](Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

[6](State Key Laboratory of Internet Architecture (Tsinghua University), Beijing 100084, China)

Abstract: Large language models (LLMs) and generative engines are reshaping how users access information on the Web. Traditional search engine optimization focuses on whether webpages appear in search results and rank highly, whereas generative engine optimization focuses on whether large language models select webpage content and incorporate it into generated answers. Existing studies have mainly

* 基金项目: 中央高校青年教师科研创新能力支持项目 (SRICSPYF-ZY2025001); 国家自然科学基金 (62441230, 62622221, 62572473, 62425201); 互联网体系结构全国重点实验室课题 (HLW2025ZD17)
收稿时间: 2026-07-29; 修改时间: 2026-09-02; 采用时间: 2026-09-13; jos 在线出版时间: 2026-11-02

focused on English-language settings, while systematic research on Chinese webpages and domestic generative platforms remains lacking. To address this gap, this study develops an evaluation framework for Chinese generative engine optimization, covering Chinese dataset construction, optimization strategy design, and source visibility assessment. The effectiveness of Chinese-language generative engine optimization (GEO) is analyzed across strategies, platforms, domains, and source positions. The study further compares GEO performance in Chinese and English settings and summarizes its practical implications for Chinese-language applications.

Key words: large language model (LLM); search engine optimization (SEO); webpage content rewriting; citation analysis; cross-platform evaluation

随着大语言模型 (large language model, LLM) 和检索增强生成 (retrieval-augmented generation, RAG) 技术的发展, 越来越多的用户开始通过生成式搜索直接获取答案[1]. 生成式搜索 (generative search) 可以检索外部信息, 并结合多个来源生成回答[2], 正在逐渐改变用户获取网络信息的方式. 与此同时, 我国也在持续推动人工智能的广泛应用: 2025 年, 国务院印发《国务院关于深入实施“人工智能+”行动的意见》, 提出推动人工智能与经济社会各行业各领域广泛深度融合[3]; 2026 年, 《中华人民共和国国民经济和社会发展第十五个五年规划纲要》进一步提出全面实施“人工智能+”行动, 培育智能原生新模式、新业态[4]. 在技术发展和应用普及的推动下, 生成式搜索正逐渐成为用户获取网络信息的重要入口.

当用户不再逐一浏览搜索结果, 而是直接获取模型生成的答案时, 网页的曝光机制也随之改变. 在传统搜索中, 网页曝光主要取决于搜索结果排名, 因此搜索引擎优化 (search engine optimization, SEO) 主要通过促进网页收录并提高自然搜索排名, 来增加其被用户访问的机会[5−7]. 生成式搜索则不同: 模型在检索网页后, 还会进一步选择来源、提取相关信息, 并据此生成回答. 因此, 网页即使已经被检索到, 也不一定会被模型采用或引用. 网页优化的目标也从提高搜索排名, 进一步扩展到提高网页内容被模型采用并呈现在生成式回答中的可能性. 生成式引擎优化 (generative engine optimization, GEO) 正是在这一背景下受到关注[8]. 如何通过优化网页内容, 提高其在生成式回答中的采用率和引用率, 已成为学术界和产业界关注的重要研究问题[8−10].

现有 GEO 研究主要基于英文问题、英文网页和国外生成式平台, 中文场景下仍缺乏系统性研究. 中文与英文在词语边界、句法结构、术语表达和篇章组织等方面存在差异[11−13], 国内生成式平台的信息来源和内容处理方式也各有特点[2,14]. 因此, 英文场景下形成的优化策略能否直接适用于中文网页, 仍有待进一步验证. 有必要在统一的中文数据和实验条件下, 系统评估不同内容表达方式对网页来源可见度的影响.

围绕上述问题, 本文构建了一套面向中文场景的生成式引擎优化评测框架, 选取多个具有代表性的领域, 收集中文问题及其相关网页, 并在保证事实一致性的前提下设计多种内容改写策略. 随后, 本文利用国内多个主流大模型系统生成带有来源标注的回答, 考察不同内容优化方式对网页可见度的影响, 并进一步从策略、平台、领域和候选来源位置等方面分析中文生成式引擎优化的效果及差异.

本文的主要工作包括以下 3 个方面.

1) 构建面向中文生成式引擎优化研究的数据与评测框架, 包括中文问题与候选网页的构建、内容优化策略的设计以及来源可见度评价方法的建立, 为后续实验提供统一的研究基础, 具体见第 3 节.

2) 在多个国内大模型平台上开展中文生成式引擎优化实验, 系统评估不同内容优化策略的有效性, 并分析模型平台、内容领域和候选来源位置等因素对优化效果的影响, 具体见第 4 节.

3) 总结中文生成式引擎优化的主要效果特征, 并与英文 GEO 的实验结果进行比较, 从语言表达、网页内容特征和模型平台等方面分析中英文场景下策略表现的差异, 具体见第 5 节.

本文提出流畅性优化、量化表达优化、来源说明增强、结构化摘要、直接引语增强、权威语气、专业术语增强、关键词强化这 8 种中文 GEO 内容改写策略 (详见第 3.2.1 节), 并围绕策略效果、平台差异、领域差异、位置差异和中英文差异这 5 个研究问题得到以下主要结论.

1) 策略效果方面, 8 种内容改写策略均取得正向提升. 关键词强化策略的效果最为显著, 综合可见度相对提升率为 25.84%; 结构化摘要策略和专业术语增强策略效果其次, 相对提升率分别为 18.79% 和 13.48%; 其余 5 种策略的提升率为 5.62%–8.95%.

2) 平台差异方面, 关键词强化策略、结构化摘要策略和专业术语增强策略在 4 个平台上均位列前 3, 不同平台的跨策略平均提升率为 8.57%–14.60%, 专业术语增强策略兼具较高收益和较低的平台间波动.

3) 领域差异方面, 关键词强化策略、结构化摘要策略和专业术语增强策略在 6 个领域中均保持较高收益, 其中仅关键词强化策略表现出显著的领域间差异, 结构化摘要策略和专业术语增强策略则具有更稳定的跨领域表现.

4) 位置差异方面, 来源 1 的基线综合可见度最高 (23.03%), 但总体提升率最低 (9.15%); 来源 4 的基线综合可见度最低 (14.71%), 总体提升率反而最高 (14.17%).

5) 中英文 GEO 差异方面, 不同策略表现出截然不同的效果. 例如, 关键词强化策略在中文 GEO 中表现为正收益 (内容贡献率相对提升 27.17%), 但在英文 GEO 中表现为负收益 (内容贡献率相对下降 8.30%).

本文第 1 节回顾传统搜索引擎优化、生成式搜索与来源引用以及生成式引擎优化的相关研究. 第 2 节介绍网页来源可见度的研究背景, 并从语言、领域和平台等方面分析中文 GEO 的研究动机. 第 3 节介绍实验数据构建、内容改写策略、来源可见度指标和评测流程. 第 4 节围绕策略效果、模型平台、应用领域和候选来源位置提出研究问题并开展实验分析. 第 5 节进一步围绕中英文 GEO 差异提出研究问题并进行对比分析. 第 6 节总结全文, 并对未来研究方向进行展望.

1 相关工作

1.1 传统搜索引擎优化

传统搜索引擎通过抓取、索引和排序, 从大规模网页集合中返回与查询相关的结果[5]. 由于用户通常优先浏览排名靠前的网页, 其点击行为具有明显的位置偏差[7,15], 因此, 传统网页优化的核心目标是提高网页在搜索结果中的自然排名, 以获得更多曝光机会. 搜索引擎优化 (SEO) 主要从网页内容和链接关系两个方面入手[6,16], 例如优化关键词、标题和正文结构、调整站内外链接等. 与此同时, 搜索系统也通过网页垃圾检测、TrustRank 等机制[17−19], 抑制关键词堆砌、链接操纵等行为对搜索结果的影响.

近年来, 一些研究开始自动修改网页内容, 在提升搜索排名的同时兼顾内容相关性和文本质量[20,21]. 另一些研究则从竞争视角出发, 分析多个内容提供者在排名机制下的策略选择, 以及不同优化行为之间的相互影响[21−23]. 这些研究仍以链接列表中的自然搜索排名为主要优化目标. 生成式搜索在检索之后还需要进行来源选择、信息提取和回答组织. 因此, 网页进入候选集合后能否继续被模型采用, 仍取决于后续的内容处理过程, 这也构成了区别于传统 SEO 的新的研究问题[8].

1.2 生成式搜索与来源引用

检索增强生成通过引入外部数据和文档, 为大语言模型生成回答提供额外的知识来源[1,24]. 针对如何更有效地获取和利用外部知识, 已有研究分别从检索和生成两个环节展开探索. 在检索环节, 密集段落检索提高了开放域问答中相关证据的召回能力[25]; 在生成环节, Fusion-in-Decoder[26]验证了模型综合多篇文档生成答案的能力. 此外, 引入外部检索还能够减少知识型对话中的事实性错误[27]. 随着生成式系统开始综合多个外部来源, 研究关注点也进一步从“能否检索到相关文档”延伸到“哪些文档最终被模型采用, 以及它们对回答产生了多大影响”[28,29].

为了便于用户核对回答依据, 生成式系统通常会在相关陈述后标注信息来源. WebGPT[30]、GopherCite[31]等工作将网页检索、回答生成和来源标注结合起来, 为带引用生成提供了较早的实践基础. 此后, 相关研究主要从两个方面评估引用质量: 一是需要引用的陈述是否给出了来源, 即引用完整性; 二是所引来源能否支持相应陈述, 即引用准确性[29,32]. 在此基础上, 归因问答研究进一步关注生成内容与可核验来源之间的对应关系, 以判断回答中的信息能否追溯到明确的证据来源[28,33,34].

已有研究表明, 实际生成式系统仍普遍存在陈述缺少来源、来源与陈述不匹配等问题[29,32,35], 类似现象在跨语言问答中同样存在[13]. 针对这些问题, 后续研究进一步区分不同类型的归因错误, 并探索证据支持判断和细粒度证据定位等方法[34,36,37]. 这类研究主要关注生成回答中的引用是否完整、准确且可核验, 而对于网页发布者而言, 还需要进一步关注自身网页能否被模型选中并用于生成回答, 以及其内容在回答中发挥了多大作用. 因此, 如

何通过调整网页内容, 提高其在生成式回答中的来源可见度[8], 成为生成式搜索中一个新的优化问题.

1.3 生成式引擎优化

生成式引擎优化 (GEO) 是指通过调整网页的内容表达和组织方式[8], 提高网页内容被生成式引擎选中、引用并用于生成回答的可能性. 与传统 SEO 主要关注网页能否进入搜索结果并获得较高排名不同, GEO 关注网页进入候选来源后, 其内容能否被生成式引擎识别、提取并组织进最终回答. 因此, 来源可见度, 即网页在生成式回答中被采用和呈现的程度, 成为衡量 GEO 效果的核心依据.

现有 GEO 研究主要考察网页内容特征与来源可见度之间的关系. 早期工作研究了权威语气、统计信息、来源说明、直接引语、流畅性、专业术语和关键词等内容改写策略[8]. 此后, 研究逐渐扩展到文档结构、信息分块等方面[38,39], 并开始利用生成式平台的反馈自动选择或组合优化策略. AutoGEO[9]根据平台对候选内容的采用情况归纳优化规则; MAGEO[40]通过多智能体协作积累和复用策略经验; AgenticGEO[41]则根据反馈动态调整改写操作; FeatGEO[42]进一步从结构、内容和语言等多个层面对网页进行表示, 并在来源可见度和内容质量之间进行多目标优化. 相关研究表明, 网页的语言表达和内容组织都会影响生成式平台的来源选择, 而不同平台对同一内容特征的响应也存在差异[9,14,42].

在单篇网页内容优化的基础上, GEO 的应用范围逐渐扩展至多查询、多参与者和完整生成链路等复杂场景. 面向查询理解, Mind Reader[43]关注问题背后的潜在信息需求; 面向多查询优化, IF-GEO[44]研究在内容预算有限的情况下, 如何协调同一网页对多个查询的优化要求. 当多个内容提供者同时参与优化时, 相关研究开始关注不同参与者之间的竞争关系. 对抗性搜索优化[10]研究分析了平台偏好对来源或产品选择的影响, C-SEO bench 及相关竞争实验[45,46]则进一步分析不同参与者之间的收益变化, 以及主题相关性和候选位置对引用结果的影响.

与此同时, GEO 评测也从单一的来源曝光, 逐步扩展到对生成流程和来源作用的综合分析. CC-GSEO-Bench[39]同时考察来源曝光、内容归因、网页结构、可信性和安全性; SAGEO Arena[47]将检索、重排序和回答生成纳入统一实验环境, 并发现有利于回答生成的改写未必能够改善前置检索结果; AgentGEO[48]则沿完整生成流程分析来源未被采用的具体环节, 包括检索、网页访问、正文解析、归因和回答生成. 跨平台研究还发现, 不同生成式系统在来源构成、内容组织、查询敏感性和跨语言稳定性等方面存在明显差异[2,14,46].

1.4 现有研究的不足

现有 GEO 研究主要围绕英文语料和国外生成式平台展开[8,9,47], 面向中文网页和国内大模型平台的系统研究仍然缺失. 中英文在词语边界、篇章组织和术语表达等方面存在明显差异[11−13], 英文场景下形成的优化策略能否适用于中文网页, 尚未得到系统验证. 此外, 已有的研究在实验语料、生成平台、策略实现和评价方法上差异较大[8,42,47], 这些实验条件上的差异容易使策略本身的作用与平台因素相互混杂, 因而难以准确判断不同策略的实际效果.

现有的引用评测研究通常以网页是否被引用及引用次数来衡量来源可见度[28,29,32,39], 但引用次数既无法反映目标网页对回答内容的实际贡献, 也无法体现相关内容在回答中所处位置的差异. 针对上述问题, 本文构建统一的中文 GEO 评测框架, 对多种内容改写策略进行受控实验, 并从来源引用和内容贡献两个维度评价网页在生成式回答中的可见度. 在此基础上, 进一步分析不同平台、内容领域和候选来源位置下的策略表现.

2 背景和动机

2.1 研究背景

传统搜索与生成式搜索中的网页信息呈现流程如图 1 所示. 在传统搜索场景中, 搜索引擎根据用户查询检索相关网页, 并按照相关性等因素对搜索结果进行排序[5,7,15]. 系统最终向用户返回由网页标题、摘要和链接组成的结果列表, 用户需要从中选择感兴趣的网页, 进入页面后自行阅读、筛选和整合信息. 因此, 网页能否进入搜索结果以及排名是否靠前, 直接影响其被用户看到和访问的机会.

相比之下, 生成式引擎在接收用户问题后, 还需要理解问题、制定检索计划并获取相关网页. 随后, 系统从不

同来源中提取与问题有关的信息, 并将其重新组织为自然语言回答[1,47,48]. 用户通常无需逐一打开网页, 便可直接获得由多个来源共同支持的回答. 在这一过程中, 进入候选来源集合并不意味着网页内容一定会呈现在最终回答中, 生成式引擎还需要在候选网页之间进行选择, 并决定从各网页中提取哪些信息以及如何将其组织进回答.

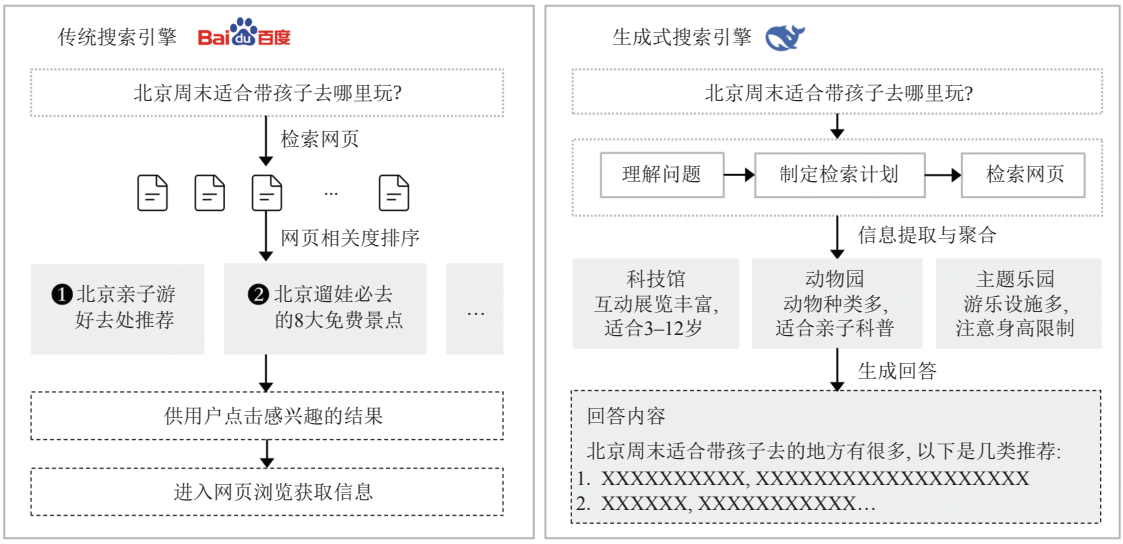


图 1 传统搜索与生成式搜索中的网页信息呈现流程

由此可见, 搜索结果排名只能反映网页在检索阶段所处的位置, 无法完整说明网页内容在生成式回答中的实际表现[8,47,48]. 排名靠前的网页如果关键信息较为分散, 或者与用户问题之间的对应关系不够明确, 模型可能只采用其中少量内容; 排名相对靠后的网页如果结论集中、表述清楚且能够直接支持回答, 也可能被模型大量采用. 因此, 在生成式搜索场景中, 网页之间的竞争已从搜索结果排名进一步延伸到回答生成阶段的来源选择、内容提取和归因呈现.

本文将网页被标注为来源及其内容被用于生成回答的程度称为来源可见度. 来源可见度主要包括两个方面: 一是目标网页在回答中被标注为信息来源的情况; 二是归因于该网页的回答内容所占篇幅及其所在位置. 前者反映网页在来源标注中的出现程度, 后者反映网页对回答内容的实际贡献. 已有研究发现, 来源引用次数与归因内容占比并不总是同步变化[8,14,39], 因此, 本文同时从来源引用和内容贡献两个维度衡量网页在生成式回答中的表现, 具体定义见第 3.3 节.

2.2 研究动机

首先, 从语言表达来看, 英文 GEO 主要基于英文问题和英文网页开展[8–10], 其结论未必能够直接适用于中文场景. 中文缺少显式词语边界, 同一概念还可能存在全称、简称和近义表达, 语序结构、篇章衔接和术语使用方式也与英文不同[11,12,49]. 这些语言特点会影响模型对问题关键词、领域概念和关键信息的识别. 因此, 英文 GEO 中围绕文本流畅性、专业术语和关键词等内容特征形成的优化方法, 仍需在中文网页和国内生成式平台上重新验证.

除了语言差异, 不同领域的网页在信息构成和表达要求上也存在明显区别. 金融投资、医疗健康和法律政务等领域通常包含风险提示、适用条件和不确定性说明, 过度突出局部信息或增强语气可能会改变原文含义, 甚至削弱必要的限制条件. 其他领域则可能更强调概念解释、操作步骤或信息组织. 为兼顾优化效果与信息质量, 本文对各类策略设置统一改写约束, 要求改写后保留原文的事实内容、结论方向、风险信息和适用范围, 具体见第 3.2.2 节.

在语言和领域差异之外, 生成式平台自身的来源选择、信息提取和回答组织方式也会影响策略效果[2,9,14]. 为此, 本文在多个国内大模型平台上开展统一条件下的对照评测, 既考察不同策略的整体效果, 也分析其跨平台表现.

3 中文生成式引擎优化

本节介绍所构建的中文生成式引擎优化评测框架, 其整体结构如图 2 所示. 该框架由 4 个相互衔接的部分组成. 首先, 中文问题集和候选网页构成评测对象, 即中文数据集构建; 其次, 中文内容改写策略生成不同版本的网页正文; 再次, 来源可见度指标用于衡量各版本网页在回答中的表现; 最后, 统一评测过程将数据、策略和指标结合起来, 通过对照实验评价不同内容改写方式的优化效果.



图 2 中文生成式引擎优化评测框架

3.1 中文数据集构建

3.1.1 中文问题集

为覆盖不同类型的中文信息需求, 本文从公开的社交媒体讨论和用户提问中收集候选问题, 并依次进行问题筛选与去重、领域归类和人工校订, 最终构建了覆盖金融投资、医疗健康、法律政务、教育学术、科技数码和生活文化社会这 6 个领域的中文问题集. 每个领域包含 100 个问题, 共计 600 个. 问题集的领域构成见表 1, 各领域的示例问题如图 3 所示. 该问题集已在 GitHub 上公开 (https://github.com/litonglab/CNGEO).

表 1 中文问题集的领域构成

领域	数量	主要覆盖内容
金融投资	100	基金、ETF、理财、资产配置、投资风险和金融知识等
医疗健康	100	疾病常识、用药、体检、健康管理和生活方式等
法律政务	100	法律咨询、政策理解、权益保障和政务办理流程等
教育学术	100	升学考试、学习方法、学术概念和教育政策等
科技数码	100	软件、硬件、人工智能、手机和计算机使用等
生活文化社会	100	旅游、消费、文化现象、公共生活和社会议题等

中文问题集的构建过程包括以下 3 个步骤.

- 步骤 1: 问题筛选与去重. 对候选问题进行语义筛选与去重, 剔除内容相同或信息需求高度重合的问题, 同时排除范围过宽、仅包含单个关键词、过度依赖即时信息或难以形成稳定答案的问题. 保留的问题应当表述自然、语义完整、信息需求明确, 并能够通过现有中文网页获得较为完整且相对稳定的回答.
- 步骤 2: 领域归类. 根据问题的主要信息需求确定领域标签. 对于同时涉及多个领域的问题, 以回答该问题所需的核心知识作为划分依据, 从而保证领域划分的一致性.
- 步骤 3: 人工校订. 对保留问题进行人工检查, 主要修正脱离原始语境后难以理解的指代、不完整表述以及

其他语言问题, 并对表述不自然或语义不清的内容进行调整, 同时保持问题原有的信息需求不变.

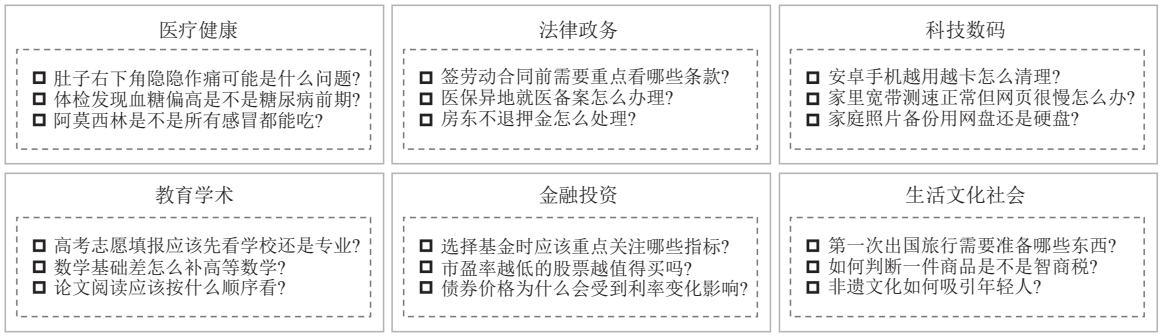


图 3　中文问题集各领域的示例问题

3.1.2　候选网页获取与正文抽取

对于每个中文问题, 本文从中文搜索引擎的自然搜索结果中构建候选网页集合, 具体包括以下两个阶段.

● 阶段 1: 候选网页筛选. 按照搜索结果的排序依次审查网页, 筛选与问题相关且正文完整可用的页面, 直至获得 k 篇有效网页. 为避免候选来源过度集中, 同一主域名最多保留两篇. 对于无法访问、链接跳转异常、正文为空或明显残缺、仅为网站首页/短视频页面, 或内容与问题无关的网页, 均予以排除, 并继续向后筛选. 入选网页按照筛选后的相对顺序依次编号为来源 1–来源 k, 形成固定候选来源序列, 其输入顺序在后续实验中保持不变.

● 阶段 2: 网页正文抽取. 对入选网页提取标题和正文, 仅保留与正文相关的主体内容, 去除评论、推荐内容及其他附属信息. 抽取过程中保留原文中的数值信息、风险提示、适用条件和免责声明, 避免因正文抽取而改变网页的事实边界和语义范围. 经过筛选与正文抽取后, 每个问题对应 k 篇候选网页, 抽取后的正文随后进入统一的中性预处理和策略改写阶段.

3.2　内容改写策略

3.2.1　策略设计

本文以英文 GEO 提出的 9 种内容优化策略为基础[8], 结合中文表达特点和事实一致性要求对策略集合进行调整. 这 9 种策略分别为: 易于理解 (easy-to-understand)、权威语气 (authoritative)、专业术语 (technical term)、流畅性优化 (fluency optimization)、来源引用 (cite source)、引语添加 (quotation addition)、统计信息添加 (statistics addition)、关键词堆叠 (keyword stuffing) 和独特词汇 (unique word).

上述易于理解与流畅性优化都涉及表达简化、语序调整和可读性提升, 操作内容存在较大重合, 因此本文将二者合并为流畅性优化. 独特词汇策略依赖对词语边界和词汇独特性的判定, 而中文缺少显式的词语分隔, 其结果容易受到分词方法和参照语料的影响, 难以形成稳定且可复现的操作规则[11,49], 因此不再单独设置. 此外, 本文根据中文表达习惯对部分策略进行调整: 统计信息添加改为量化表达优化, 来源引用改为来源说明增强, 引语添加改为直接引语增强, 关键词堆叠改为关键词强化. 为进一步考察信息组织方式对来源可见度的影响[38], 本文增设结构化摘要策略. 为保证策略实现的一致性和可复现性, 本文明确规定各策略的操作目标、具体方式和适用边界.

● 流畅性优化: 以语序、句间衔接和重复表达为主要操作对象, 在保留原文信息量、事实内容和主要段落层次的基础上, 通过调整句式、合并重复内容和改善段落衔接提高文本的自然度与可读性, 不改变原文内容重点、语气强度和事实边界.

● 量化表达优化: 以原文已有的数值、时间、比例、金额、规模、频率和排名等量化信息为主要操作对象, 通过归并、前置、分组和重排提高这些量化信息的显著性, 同时保留相关事实和适用条件, 不新增原文未提供的量化信息, 不对原有量化信息进行推断、换算或修改, 也不因突出量化信息而压缩其他正文内容.

● 来源说明增强: 以原文已有的机构、作者身份、法规、报告、公告、平台和发布时间等来源线索为主要操作对象, 通过整合分散的来源信息并将其置于相关事实附近, 增强来源与内容之间的对应关系; 来源信息不足时不

主动补充外部出处.

● 结构化摘要: 以正文的信息组织方式为主要操作对象, 将原文重组为“结论先行、分层组织和要点化表达”的结构化正文, 并通过纯文本小标题集中呈现关键信息; 改写仍完整保留原文主要事实、例子、条件和风险提示, 不将正文压缩为短摘要.

● 直接引语增强: 在完整保留原文主要信息的基础上, 从原文中选取少量连续出现且语义完整的短句作为直接引语, 并嵌入相应正文段落, 以提高内容的可直接引用性; 所有引语必须能够在原文中逐字对应, 不生成、改写或拼接原文不存在的引语.

● 权威语气: 以文本措辞和陈述方式为主要操作对象, 在保持原文事实内容、论证顺序、具体信息和风险边界的基础上, 将口语化、情绪化或营销化表达调整为更加正式、准确和克制的专业表述; 不新增权威背书, 也不提高原文结论的确定程度.

● 专业术语增强: 以原文已有的领域概念为主要操作对象, 根据所属领域将相关表达规范为更加准确的专业术语, 并在必要时保留相应的通俗解释; 仅对原文已经涉及的概念进行术语化, 不引入新的概念体系、事实或结论.

● 关键词强化: 以用户问题和网页标题中的核心关键词及其近义表达为主要操作对象, 在段首、小标题、过渡句和结尾等位置提高其出现频率; 关键词强化仅围绕原文已有事实展开, 允许较高程度的重复, 但仍保持正文基本通顺, 并避免无意义的孤立词堆积.

为直观展示不同内容改写策略的具体作用方式, 图 4 给出了不同内容改写策略的局部改写示例. 不同策略分别从语言表达、量化信息、来源说明、内容结构、直接引语、语气、专业术语和关键词等方面调整网页内容, 同时遵循第 3.2.2 节规定的统一改写约束.

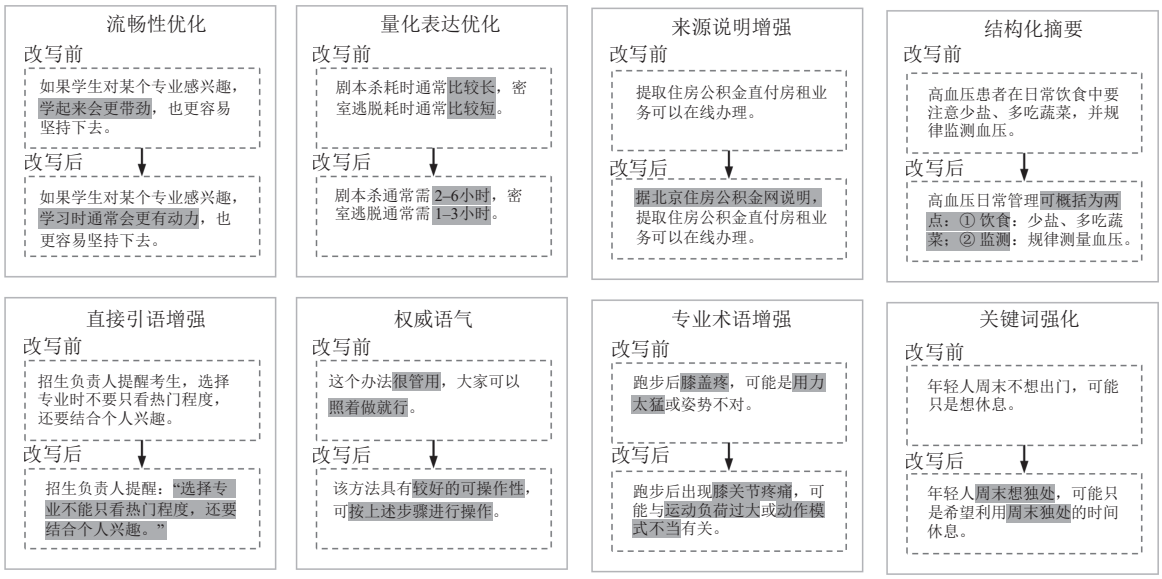


图 4 GEO 策略驱动的网页内容改写示例

3.2.2 统一改写约束

为保证不同内容改写策略之间的可比性, 本文采用两层控制方式: 策略特定指令和统一改写约束. 统一改写约束适用于全部策略, 各策略仅在其允许范围内调整目标内容特征. 统一约束主要包括以下 4 个方面.

1) 事实与证据边界. 改写只能使用原文已经明确支持的信息, 不得新增原文无法支持的事实、案例、因果关系、判断或建议, 也不得虚构统计数据、引用来源、专家观点、法规依据、报告名称、发布日期、直接引语或人物身份. 若原文信息不足以执行某项策略, 仅允许在已有事实范围内进行最小调整.

2) 语义与风险边界. 改写不得改变原文的事实内容、确定程度和适用范围, 不得将“可能”“通常”“建议”“需要

结合情况”等条件性表达改写为确定性承诺. 对于医疗、法律、金融、教育升学等通常包含较多风险提示、适用条件和限制性信息的内容, 必须完整保留原文中的风险提示、免责声明、适用条件、限制条件和专业判断要求.

3) 信息完整性与长度约束. 本文研究的是网页内容改写而非摘要生成, 因此除明显网页噪声、重复广告、无关导航和乱码外, 应保留原文的主要信息层次、关键例子、条件说明、建议和结论. 改写后的正文长度原则上控制在原文的 80%–120% 之间; 当策略需要增加少量结构或说明时, 最大不超过原文的 130%.

4) 输出格式约束. 输出必须为可直接作为网页正文使用的中文文本, 不包含改写说明、分析过程、执行清单、JSON 外壳、Markdown 代码块或其他元话语. 允许使用纯文本小标题和自然段, 但不使用与网页正文无关的格式化装饰.

以上约束中, 量化表达优化、来源说明增强和直接引语增强最容易引入原文不存在的信息, 因此进一步限制为仅对原文已有的数字、来源线索或可直接引用的表述进行重组、突出和抽取, 不得通过补充外部信息来实现策略目标.

3.2.3 改写生成与质量控制

网页正文抽取完成后, 本文首先使用 DeepSeek-v4-pro 对所有网页执行统一的中性预处理. 对于问题 q 的第 i 篇原始网页正文 $x_{q,i}$, 设 $T_0(\cdot)$ 表示统一的中性预处理操作, 则第 i 篇网页的中性改写版本表示为:

$$b_{q,i} = T_0(x_{q,i}) \tag{1}$$

中性预处理仅执行残留非正文信息清理、语句整理、段落衔接优化和格式统一, 不突出任何特定内容特征, 也不采用后续 GEO 策略. 每个问题对应的 k 篇网页均首先生成中性改写版本, 并共同构成后续评测的基线来源.

在获得中性改写版本后, 本文进一步针对每篇网页生成 8 种策略改写版本. 沿用后文的记号, 设 $T_g(\cdot)$ 表示策略 g 对应的改写操作, 则第 i 篇网页应用策略 g 后的版本记为 $T_g(b_{q,i})$. 实际调用时, 策略改写并非仅输入中性正文, 而是同时提供当前中文问题、所属领域、网页标题和待改写的中性正文; 对于同一问题和网页, 上述上下文保持不变, 因此在后续实验的形式化表示中, 将其统一抽象为操作 $T_g(\cdot)$.

为保证不同策略的实现方式一致且可复现, 本文采用模块化方式构造策略改写 Prompt. 每个策略的完整 Prompt 由以下 3 部分组成.

1) 任务输入. 包括当前中文问题、所属领域、网页标题和待改写的中性正文, 用于明确用户信息需求、领域语境及可操作的文本范围.

2) 策略特定指令. 对应第 3.2.1 节定义的具体策略, 进一步给出该策略的核心目标、执行要求、禁止事项和输出形式, 用于规定当前允许调整或强化的内容特征.

3) 统一改写约束. 加入第 3.2.2 节规定的共享约束, 对事实与证据边界、语义与风险边界、信息完整性与长度以及输出格式进行统一限制. 在生成最终正文前, Prompt 还要求模型在内部依次识别原文的核心事实、条件、例子、风险提示和结论, 判断当前策略允许调整或强化的内容, 并且仅在允许范围内重组语言.

为直观展示上述 Prompt 的具体构造方式, 图 5 以结构化摘要策略为例, 展示其 Prompt 的模块化结构和关键内容. 其他 7 类策略采用相同的整体构造方式, 仅替换相应的策略特定指令. 本文使用的中性预处理 Prompt、8 种策略改写 Prompt 和完整的统一改写约束已随数据和代码在公开仓库 (https://github.com/litonglab/CNGEO) 中提供.

中性改写和 8 种策略改写均由同一改写模型生成, 具体模型版本和生成参数见第 4.1.2 节. 每篇网页在每种改写条件下仅生成一个固定版本, 同一版本随后统一用于各生成式平台的回答生成实验, 保证不同平台使用完全一致的网页版本.

改写生成完成后, 本文按照统一改写约束对结果进行质量检查, 检查内容主要包括以下 3 个方面.

1) 格式与长度检查. 检查输出是否仅包含可直接使用的网页正文, 并验证正文长度是否满足第 3.2.2 节规定的篇幅范围.

2) 事实与语义一致性检查. 核对关键事实、数值、来源信息、结论方向、风险提示和适用条件是否得到保留, 并检查是否引入原文无法支持的新信息.

3) 策略边界检查. 根据当前策略允许的操作和禁止事项检查改写结果. 对于量化表达优化、来源说明增强和直接引语增强, 重点核对其中的数字、来源和引语是否能够由原文直接支持; 对于医疗健康、法律政务和金融投资等领域的文本, 进一步抽样复核风险边界、限制条件和不确定性表达是否得到保留.

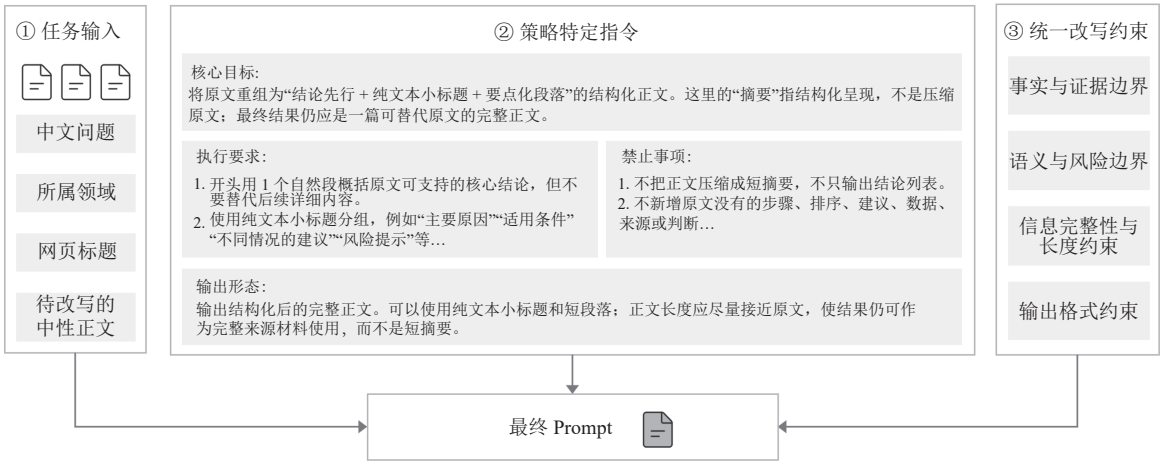


图 5 结构化摘要策略的 Prompt 构造示例

经上述检查后, 符合要求的改写版本将被固定保存, 并作为后续回答生成实验的统一网页输入.

上述检查主要用于保证改写结果满足预先规定的格式和内容约束. 为进一步独立验证改写文本的实际质量, 并检验单一改写模型及固定生成版本是否影响主要实验结论, 本文进一步开展人工改写质量评审和替代改写模型稳健性实验. 两项补充实验的具体设置见第 4.1.4 节, 实验结果见第 4.7 节.

3.3 来源可见度指标

引用次数能够反映目标网页在来源标注中出现的频率, 但不能完整表示其对回答内容的贡献. 同一网页可能被多次用于支持简短陈述, 也可能只被标注一次, 却支撑了回答中的核心结论.

除了内容篇幅, 归因内容在回答中的位置也会影响网页在回答中的可见度. 回答前部的内容通常更接近核心结论, 而后部内容更多承担背景补充和条件说明[8,50]. 此外, 同一句话还可能同时标注多个来源, 此时需要在多个来源之间分配该句的内容贡献. 因此, GEO 评测既需要考察目标网页在候选来源中的引用份额, 也需要衡量归因内容的篇幅和位置[8,29,32].

基于上述考虑, 本文采用来源引用率和内容贡献率两个指标. 来源引用率 (citation share, *CS*) 表示目标网页的引用次数在全部来源标注次数中的占比, 内容贡献率 (position-aware weighted share, *PAWS*) 根据归因内容的篇幅和所在位置衡量目标网页对回答内容的贡献. 本文进一步对 *CS* 和 *PAWS* 进行加权, 得到综合可见度, 用于比较目标网页改写前后的整体变化.

3.3.1 来源引用率

对于网页发布者而言, 仅判断目标网页是否被引用, 还无法反映其在多个候选来源中的相对表现. 为此, 本文使用来源引用率衡量目标网页在回答全部来源标注中所占的份额. 设回答中来源 i 的标注出现次数为 c_i, 则其来源引用率定义为:

$$CS(i)=\frac{c_i}{\sum_{n=1}^{k} c_n} \tag{2}$$

其中, $CS(i)$ 的取值范围为 $[0,1]$. 当回答中未出现有效来源标注时, 各来源的 *CS* 均记为 0. *CS* 越高, 说明目标网页在全部来源标注中所占的份额越大.

3.3.2 内容贡献率

来源引用率反映了目标网页在来源标注中的份额, 但仍无法说明有多少回答内容实际来自该网页. 为此, 本文进一步结合归因内容的篇幅及其所在位置计算内容贡献率. 例如, 一个来源可能被多次用于支持简短陈述, 也可能只被标注一次, 却支撑了篇幅较长且位于回答前部的内容. 将回答切分为由 N 个句子构成的有序序列 $\mathcal{A}=(s_0, s_1,\ldots,s_{N-1})$. 第 j 个句子去除来源标注和空白字符后的中文字符数记为 l_j, 该句子标注的来源编号集合记为 C_j. 当一个句子只标注一个来源时, 其全部字符计入该来源; 当一个句子同时标注多个来源时, 其字符数在 C_j 包含的各来源之间平均分配.

为体现归因内容在回答中的位置[8], 第 j 个句子的权重设为 $w_j=\exp\left(-\frac{j}{N}\right)$, 句子越靠前, 对应的权重越高. 来源 i 的内容贡献率定义为:

$$PAWS(i)=\frac{\sum\limits_{\substack{0\leqslant j<N\\ i\in C_j}}\frac{l_j}{|C_j|}w_j}{\sum\limits_{j=0}^{N-1}l_jw_j} \tag{3}$$

其中, $|C_j|$ 表示第 j 个句子标注的来源数量. 未标注来源的句子不计入分子, 但其字符数仍计入分母. 这样可以使 *PAWS* 同时反映目标来源所贡献的内容篇幅、内容所在位置以及回答中未归因内容所占的比例. $PAWS(i)$ 的取值范围为 $[0,1]$. *PAWS* 越高, 说明归因于目标网页的回答内容越多, 或者相关内容在回答中的位置越靠前.

3.3.3 综合可见度

CS 反映目标网页在来源标注中的份额, *PAWS* 反映归因于该网页的回答内容篇幅和位置. 为了同时考虑两个方面, 本文将二者加权得到来源 i 的综合可见度:

$$V(i)=\alpha\cdot PAWS(i)+(1-\alpha)\cdot CS(i) \tag{4}$$

其中, $\alpha\in[0,1]$ 用于控制 *PAWS* 和 *CS* 的相对权重. 综合可见度的具体权重取值在第 4 节实验设置中说明, 并通过稳健性分析检验权重变化对策略排序的影响.

来源可见度指标的计算过程如图 6 所示. 系统首先识别回答中的来源标注并完成句子切分, 随后分别计算来源引用率、内容贡献率和综合可见度.

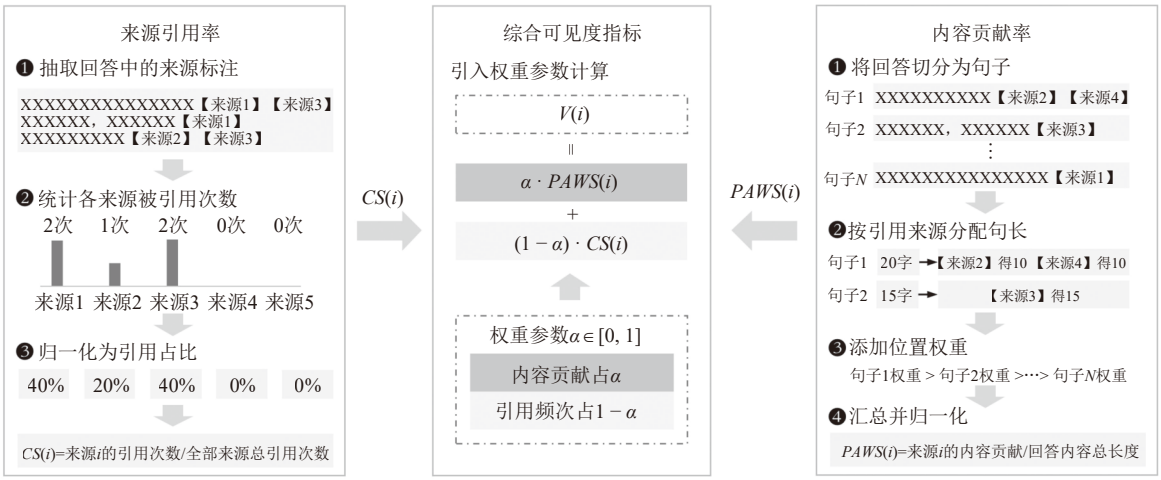


图 6 来源可见度指标计算框架

3.4 中文 GEO 评测过程

第 3.1–3.3 节分别介绍了中文数据集、内容改写策略和来源可见度指标, 本节进一步说明上述要素如何组成

完整的中文 GEO 评测过程. 该过程通过控制候选来源、来源顺序和回答要求, 比较目标网页在中性改写和策略改写条件下的来源可见度变化. 本文重点考察候选来源确定后的回答生成阶段, 通过固定候选来源及其输入顺序, 排除其他因素的干扰, 单独考察网页内容表达变化的影响. 该受控设置与英文 GEO 的评测思路一致[8].

3.4.1 控制变量设定

为单独考察网页内容改写对回答生成的影响, 本文固定每轮实验中的候选来源及其输入顺序, 并将大模型平台视为根据问题和给定来源生成回答的黑盒函数. 沿用第 3.2.3 节中的记号, 对于问题 q, 其 k 篇完成正文抽取的网页构成有序序列 $\mathcal{X}_q=(x_{q,1},x_{q,2},\ldots,x_{q,k})$, 其中, $x_{q,i}$ 表示候选序列中位于第 i 位的网页. 各网页经统一中性预处理后得到 $b_{q,i}$, 因此基线条件下的候选来源序列表示为 $\mathcal{B}_q=(b_{q,1},b_{q,2},\ldots,b_{q,k})$.

当第 i 篇网页作为目标来源并应用策略 g 时, 策略条件下的来源序列为 $\mathcal{B}_q^{(g,i)}=(b_{q,1},b_{q,2},\ldots,T_g(b_{q,i}),\ldots,b_{q,k})$, 该序列仅将目标网页的中性改写版本替换为相应的策略改写版本, 其余 $k-1$ 篇网页及其输入位置均保持不变.

设平台 p 根据问题和给定来源生成回答的函数为 $F_p(q,\mathcal{B})$, 则基线回答和策略回答分别表示为 $A_{q,p}^{(0)}=F_p(q,\mathcal{B}_q)$ 和 $A_{q,p}^{(g,i)}=F_p\left(q,\mathcal{B}_q^{(g,i)}\right)$. 通过比较两类回答中目标网页的来源可见度, 可以衡量策略 g 对目标网页表现的影响.

3.4.2 统一回答提示词

各生成式平台采用相同的回答提示. 提示要求模型严格依据给定来源回答问题, 并在相关陈述后使用【来源 i】标注回答依据. 每次回答生成均输入一个中文问题和 k 篇候选网页. 每篇来源包括网页标题和当前实验条件下的正文版本, 即基线条件输入中性改写正文, 策略条件仅将目标来源替换为相应的策略改写正文.

各来源按照固定顺序编号为来源 1–k. 模型需要使用自然、流畅的中文直接作答, 每个包含事实、观点、建议、判断或结论的句子均应在句末标注来源; 当同一陈述由多个来源共同支持时, 可以连续标注多个来源.

对于与问题关联较弱、内容重复或无法支持回答的来源, 模型可以不予采用, 但不得补充给定来源之外的信息. 回答提示与输入示例如图 7 所示. 为便于展示, 图 7 中仅保留各来源的标题和部分正文, 正式实验使用完整的内容版本.

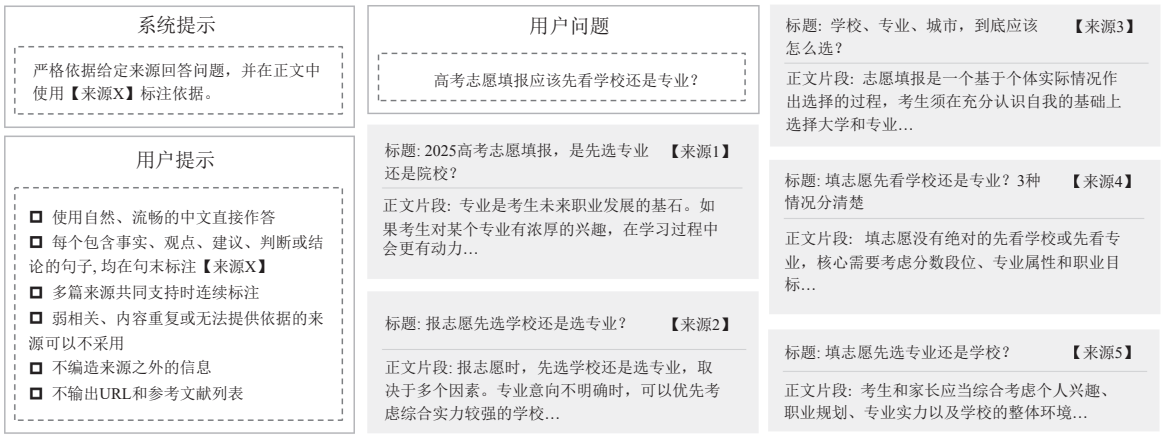


图 7 统一回答提示与输入示例

回答生成后, 系统识别并解析其中的来源标注. 解析过程同时支持单来源标注、多个来源连续标注, 以及同一来源在回答中多次出现的情况. 回答按照中文句号、问号、感叹号和分号进行切分, 并建立各句内容与来源编号之间的对应关系.

在此基础上, 系统计算每个目标网页的来源引用率、内容贡献率和综合可见度, 并与相应基线结果进行配对比较. 实验结果进一步按照内容改写策略、生成式平台、问题领域和候选来源位置进行汇总, 用于分析不同因素对中文网页来源可见度的影响.

4 实验与评估

本节围绕中文 GEO 的策略效果、大模型平台、应用领域和候选来源位置展开实验分析, 重点回答以下 4 个研究问题 (research question, RQ).

- RQ1 (策略效果): 不同内容改写策略的来源可见度提升效果有何差异?
- RQ2 (平台差异): 不同大模型平台下各策略的提升效果与稳定性有何差异?
- RQ3 (领域差异): 不同应用领域下各策略的提升效果是否存在显著差异?
- RQ4 (位置差异): 不同候选来源位置下的基线可见度与策略提升空间有何差异?

第 4.1 节介绍实验设置. 第 4.2–4.5 节分别针对 RQ1–RQ4 展开分析. 第 4.6 节进一步检验综合可见度权重设置对实验结论的影响. 第 4.7 节从人工质量评审和替代改写模型两方面验证内容改写过程的可靠性与结论的稳健性.

4.1 实验设置

4.1.1 实验数据与任务规模

本实验覆盖金融投资、医疗健康、法律政务、教育学术、科技数码和生活文化社会这 6 个领域, 共包含 600 个中文问题, 中文搜索引擎选用百度. 第 3 节将每个问题对应的候选网页数量统一记为 k, 本实验取 $k=5$, 因此共包含 3 000 篇候选网页. 实验设置 8 种内容改写策略, 并选取豆包[51]、DeepSeek[52]、文心一言[53]和通义千问[54]这 4 个国内大模型平台进行评测. 为降低单次生成波动对实验结果的影响, 每种生成条件均重复运行 5 次.

在基线条件下, 每个问题对应的 5 篇候选网页均使用中性改写版本. 每个问题分别在 4 个平台上生成回答, 并重复 5 次, 因此共得到 12 000 条基线回答; 在策略条件下, 5 篇候选网页依次作为目标来源, 每次仅将目标网页的中性改写版本替换为相应的策略改写版本, 其余 4 篇网页仍保持中性改写版本, 候选来源数量及输入顺序均不变. 每个目标来源分别应用 8 种内容改写策略, 并在 4 个平台上重复生成 5 次, 共得到 480 000 条策略回答. 实验最终生成 492 000 条回答.

4.1.2 实验平台与调用配置

回答生成实验选用 DeepSeek、通义千问、豆包和文心一言这 4 个平台, 分别调用 DeepSeek-v4-flash、Qwen3.6-flash-2026-04-16、Doubao-seed-2-1-turbo-260628 和 ERNIE-4.5-turbo-20260402 模型[51–54]. 为保证不同平台之间实验条件的一致性, 4 个平台均关闭思考模式, 最大输出长度统一设为 4 000 tokens, Temperature 和 Top-p 均设为 1.0. 实验过程中, 各平台使用相同的问题、候选来源及其输入顺序, 并采用统一的回答提示. 每种生成条件重复运行 5 次, 以降低单次生成波动对实验结果的影响.

网页正文抽取完成后, 首先使用 DeepSeek-v4-pro 对全部网页进行统一的中性预处理, 包括内容清洗、语句整理、段落衔接优化和格式统一, 并生成中性改写版本. 8 种策略改写均以中性改写版本为输入, 并由同一模型生成. 为减少改写过程中的随机差异, 模型关闭思考模式, 最大输出长度设为 8 192 tokens, Temperature 设为 0, Top-p 设为 1.0. 每篇网页在每种改写条件下仅生成一个版本, 同一版本随后统一用于 4 个平台的回答生成实验, 以避免改写版本差异对平台比较结果产生干扰. 主实验采用固定单一改写版本以保证 4 个平台之间输入内容完全一致; 考虑到该设置可能使结果依赖特定改写模型和单次生成结果, 本文另外通过第 4.1.4 节所述替代改写模型实验进行稳健性验证.

4.1.3 评价指标与统计方法

本文按照第 3.3 节的定义, 采用来源引用率 CS、内容贡献率 $PAWS$ 和综合可见度 V 评价目标来源的表现. 主实验中, 综合可见度的权重参数 α 取 0.7, 即 $PAWS$ 和 CS 的权重分别为 70% 和 30%, 使综合指标更侧重目标网页对回答内容的实际贡献. 3 项指标的取值范围均为 0–1, 正文、表格和图中统一换算为百分数. 对于每个实验条件, 本文先分别计算 5 次重复生成结果的指标值, 再取平均值作为该条件的最终结果.

本文从绝对变化和相对变化两个方面衡量策略效果. 绝对变化指策略条件与基线条件之间的平均指标差, 单位为百分点; 相对变化指分组内策略条件相对于基线条件的提升比例, 单位为百分比. 由于部分样本在基线条件下的可见度为 0, 无法计算单个样本的相对变化, 本文先分别汇总分组内的策略结果和基线结果, 再根据二者的差额

与基线汇总值计算相对提升率. 整体、平台、领域和候选来源位置等分析均采用这一计算方法, *CS* 和 *PAWS* 的相对提升率也按相同口径计算.

对于各策略的整体效果, 本文采用按领域分层、以问题为簇的 Bootstrap 方法[55]估计相对提升率的置信区间. 每次重采样分别在 6 个领域内对 100 个问题进行有放回抽样, 对于每个被抽中的问题, 保留其在 4 个平台、5 个候选来源位置以及基线和策略条件下的全部平均结果, 以维持同一问题下各项观测之间的相关性. 本文共进行 2000 次重采样, 并以相对提升率分布的 2.5% 和 97.5% 分位数构造 95% 置信区间.

为进一步检验不同应用领域下策略效果是否存在统计显著差异, 本文以问题为统计单位构造问题级可见度变化. 对于每个问题和每种策略, 首先按照上述口径对 5 次重复生成结果取平均, 再对 4 个生成式平台和 5 个候选来源位置进行汇总, 计算策略条件与对应基线条件之间的综合可见度绝对变化 ΔV. 由此, 每个问题在每种策略下仅保留 1 个问题级观测值. 随后, 将 600 个问题按照 6 个领域划分为 6 组, 每组 100 个问题, 采用 Kruskal-Wallis 检验比较各领域的问题级 ΔV 分布. 本文分别对 8 种内容改写策略及其总体结果进行检验, 其中, 8 种策略总体结果定义为同一问题在 8 种策略下问题级 ΔV 的算术平均值, 共得到 9 组检验结果, 并采用 Benjamini-Hochberg 方法进行多重检验校正, 显著性水平设为 0.05.

4.1.4 改写质量与稳健性验证设置

为验证中性预处理和策略改写文本的实际质量, 本文从 600 个中文问题中按照领域进行分层抽样, 每个领域随机抽取 10 个问题, 共 60 个问题. 对于每个问题选取 1 篇网页作为人工评审对象, 并在每个领域内均匀覆盖来源 1–来源 5 这 5 个候选位置. 由此构造 60 组“原始网页-中性预处理文本”对, 以及 480 组“中性预处理文本-策略改写文本”对, 共 540 组文本对.

两名评审者在不查看对方评分的情况下独立进行 1–5 分评价, 分数越高表示相应要求的完成程度越高. 对于中性预处理任务, 分别评价事实一致性、预处理任务完成度和文本质量; 对于策略改写任务, 分别评价事实一致性、策略实现程度和文本质量. 其中, 事实一致性主要考察改写文本是否保留原文事实、结论方向、适用条件和风险信息, 并避免引入原文无法支持的新信息; 预处理任务完成度主要考察文本是否完成非正文信息清理、语句整理、段落衔接优化和格式统一, 同时未突出特定内容特征; 策略实现程度主要考察改写结果是否按照第 3.2.1 节的定义实现对应策略; 文本质量主要考察改写文本的完整性、自然度、流畅性以及是否存在明显的语言和结构问题. 评审时向评审者提供相应的待比较文本及预处理任务或策略定义, 但不提供相应文本的来源可见度实验结果, 以避免实验结果对质量判断产生影响. 本文进一步统计两名评审者评分的平均值、分差不超过 1 分的样本比例以及平均绝对分差.

为检验主要策略结论是否依赖单一改写模型及其固定生成版本, 本文在上述 60 个问题组成的稳健性子集上使用另一改写模型重复实验. 对于每个问题的 5 篇候选网页, 使用 Qwen-3.8-max 模型重新进行中性预处理, 并在此基础上分别生成 8 种策略改写版本, 共涉及 300 篇候选网页. 回答生成阶段仍采用与主实验完全相同的 4 个生成式平台、回答提示、候选来源顺序和生成参数, 每种条件重复生成 5 次. 基线条件下, 5 篇候选网页全部采用 Qwen-3.8-max 生成的中性预处理版本; 策略条件下, 仅将目标来源替换为对应的 Qwen-3.8-max 策略改写版本, 其余 4 篇仍使用 Qwen-3.8-max 生成的中性预处理版本, 从而保证同一实验条件下的候选网页均来自相同改写模型. 最后, 将该 60 个问题子集上的 DeepSeek-v4-pro 结果与 Qwen-3.8-max 结果进行比较, 并通过策略提升方向、策略排序以及 Spearman 等级相关系数评价主要实验结论的一致性.

4.2 中文 GEO 策略的整体效果与分项指标分析

针对 RQ1, 本节比较 8 种内容改写策略在全部平台、领域和候选来源位置上的整体效果, 并进一步分析不同策略在来源引用率和内容贡献率上的表现. 表 2 给出了 8 种策略在全部平台、领域和候选来源位置上的整体结果. 基线条件下, 来源引用率 *CS*、内容贡献率 *PAWS* 和综合可见度 *V* 分别为 20.00%、16.76% 和 17.73%. 由于每个问题包含 5 篇候选网页, 且 5 篇网页依次作为目标来源, 基线 *CS* 的整体均值为 20.00%, 与 5 来源条件下的归一化结果一致.

表 2　不同 GEO 策略的整体实验结果

策略	*CS* (%)	*PAWS* (%)	*V* (%)	可见度差值 (%)	相对提升率 (%)	95%置信区间
基线	20.00	16.76	17.73	—	—	—
关键词强化	24.65	21.32	22.32	4.58	25.84	[23.84, 27.78]
结构化摘要	23.34	20.09	21.07	3.33	18.79	[17.42, 20.15]
专业术语增强	22.33	19.18	20.13	2.39	13.48	[12.44, 14.57]
直接引语增强	21.62	18.34	19.32	1.59	8.95	[8.01, 9.99]
权威语气	21.61	18.31	19.30	1.57	8.83	[7.89, 9.83]
量化表达优化	21.27	17.88	18.90	1.16	6.56	[5.61, 7.52]
流畅性优化	21.20	17.81	18.82	1.09	6.14	[5.14, 7.08]
来源说明增强	21.13	17.70	18.73	1.00	5.62	[4.76, 6.52]

8 种策略的综合可见度均高于基线, 其相对提升率的 95% Bootstrap 置信区间均未包含 0, 说明网页内容经过有针对性的改写后, 更容易被模型选用并纳入回答. 不过, 不同策略之间的提升幅度存在明显差异. 关键词强化的综合可见度由 17.73% 提高至 22.32%, 结构化摘要和专业术语增强分别达到 21.07% 和 20.13%, 明显高于其余策略. 由此可见, 提高网页与问题之间的语义对应程度, 并降低模型识别和提取关键信息的难度, 是中文 GEO 最有效的优化方向; 而单纯改善语言质量或增强内容权威感的收益相对有限.

从具体作用方式来看, 3 种主要策略分别改善了网页与问题的对应关系、关键信息的组织方式和领域概念的表达. 关键词强化使网页对问题核心概念的覆盖更加直接, 降低了模型识别网页相关性的难度; 结构化摘要将结论和关键依据集中呈现, 提高了内容被抽取和重新组织的效率; 专业术语增强则使网页中的概念表达更加规范, 有助于模型更稳定地识别和整合相关信息. 三者共同说明, 生成式平台更偏好主题明确、信息集中且概念表达稳定的网页内容. 相比单纯改善语言表达, 降低模型识别和提取关键信息的难度更有助于提高来源可见度.

各策略在综合可见度、*CS* 和 *PAWS* 上的相对提升结果见图 8. 直接引语增强和权威语气的综合可见度相对提升率接近 9%, 量化表达优化、流畅性优化和来源说明增强则介于 5.62%–6.56% 之间. 后几类策略虽然能够改善局部表达, 但没有明显增强网页与问题之间的对应关系, 也没有重新组织回答所需的核心信息, 因此提升幅度较小.

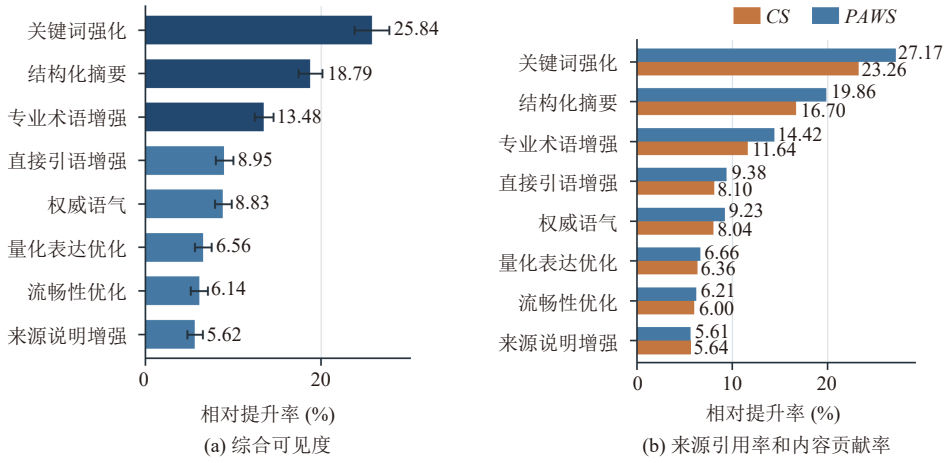


图 8　不同 GEO 策略的相对提升率

从 *CS* 和 *PAWS* 的分项结果看, 关键词强化、结构化摘要和专业术语增强对 *PAWS* 的提升均高于对 *CS* 的提升. 3 种策略对 *PAWS* 和 *CS* 的提升差值分别为 3.91、3.16 和 2.78 个百分点. 这表明, 它们带来的收益不只是增加目标网页的引用次数, 而是使更多回答内容来自该网页. 换言之, 有效的 GEO 优化并不是单纯提高网页被引用的概率, 而是提高网页在回答内容中的实际占比.

相比之下, 来源说明增强对 *CS* 和 *PAWS* 的相对提升率分别为 5.64% 和 5.61%, 说明该策略对来源引用率和内

容贡献率的提升幅度较小, 且两者基本同步. 结合 8 种策略的整体结果, 中文网页的优化优先级可以概括为: 首先强化网页与问题之间的关键词和概念对应, 其次集中呈现结论和关键依据, 再通过规范术语提高内容的可识别性; 流畅性、来源说明和语气调整更适合作为辅助优化.

4.3 平台差异与跨平台稳定性

针对 RQ2, 本节比较 8 种内容改写策略在 4 个国内大模型平台上的提升效果及跨平台波动. 各策略在 4 个平台上的相对提升率如图 9 所示. 关键词强化、结构化摘要和专业术语增强在各平台上均位列前 3, 所有策略与平台的组合也均取得正向提升. 由此可以看出, 中文网页中词汇对齐、信息集中和术语规范这 3 类特征, 并非仅对某一平台有效, 而是在多个国内生成式平台上表现出较为一致的优势. 这说明上述策略具有较好的跨平台适用性, 有助于模型更准确地识别网页相关性并提取关键信息.

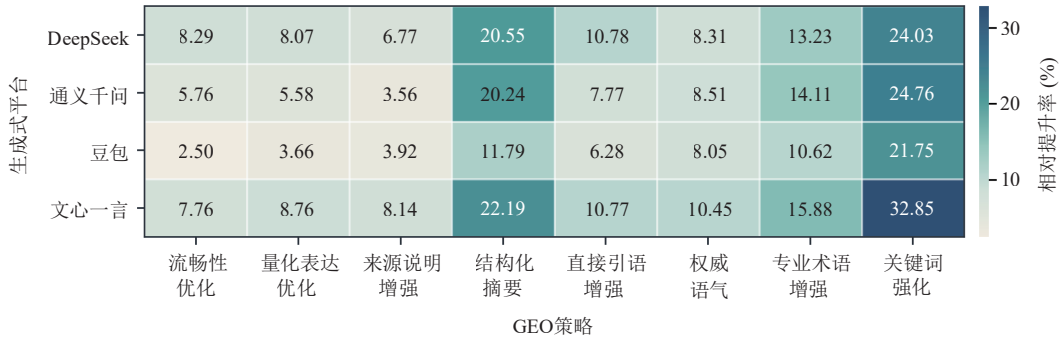


图 9 不同大模型平台下 GEO 策略的总体相对提升率

不过, 各平台对内容改写的响应幅度存在明显差异. 对 8 种策略的相对提升率取平均后, 文心一言、DeepSeek、通义千问和豆包的跨策略平均提升率分别为 14.60%、12.50%、11.29% 和 8.57%, 最高与最低相差 6.03 个百分点. 关键词强化在文心一言上的相对提升率达到 32.85%, 在豆包上则为 21.75%; 结构化摘要在两个平台上的提升率分别为 22.19% 和 11.79%. 这意味着, 同一种内容优化并不会在不同平台上产生相同的收益, 平台特性会影响网页改写能够带来的提升幅度. 例如, 文心一言的来源可见度对内容表达变化更为敏感, 豆包对内容改写的响应相对较弱, 单纯改写网页内容所能带来的提升相对有限.

各策略的跨平台平均效果及其波动如图 10 所示. 关键词强化和结构化摘要的平均提升最高, 但平台间标准差也分别达到 4.20 和 4.06 个百分点. 这两种策略具有较高的收益上限, 且在不同平台上的作用方向保持一致, 但具体提升幅度更容易受到平台差异影响.

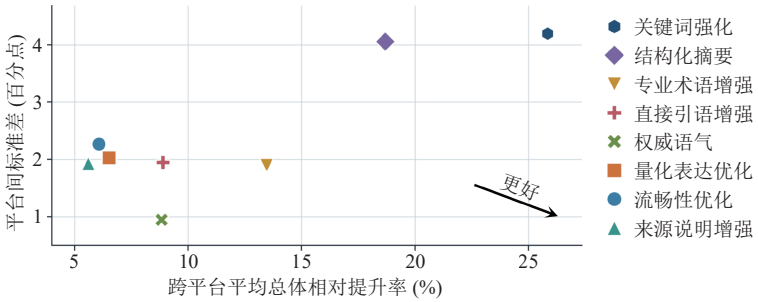


图 10 GEO 策略的跨平台效果与稳定性

综合来看, 中文 GEO 应当区分通用优化与平台适配两个层次. 专业术语增强可作为跨平台优化的基础策略, 关键词强化和结构化摘要则用于进一步提高来源可见度. 对于只面向单一平台的网页, 应根据目标平台的响应特征调整策略重点; 对于需要同时覆盖多个平台的内容, 不宜过度依赖平台间波动较大的单一策略, 而应先保证术语规范和信息清晰, 再有针对性地加强关键词对应和信息组织.

4.4 不同应用领域的策略效果与显著性分析

针对 RQ3, 本节比较 8 种内容改写策略在 6 个应用领域中的效果差异. 8 种策略在 6 个应用领域中的相对提升率如图 11 所示. 关键词强化、结构化摘要和专业术语增强在各领域中均位列前 3, 表明这 3 类优化在不同领域中均保持较高收益.

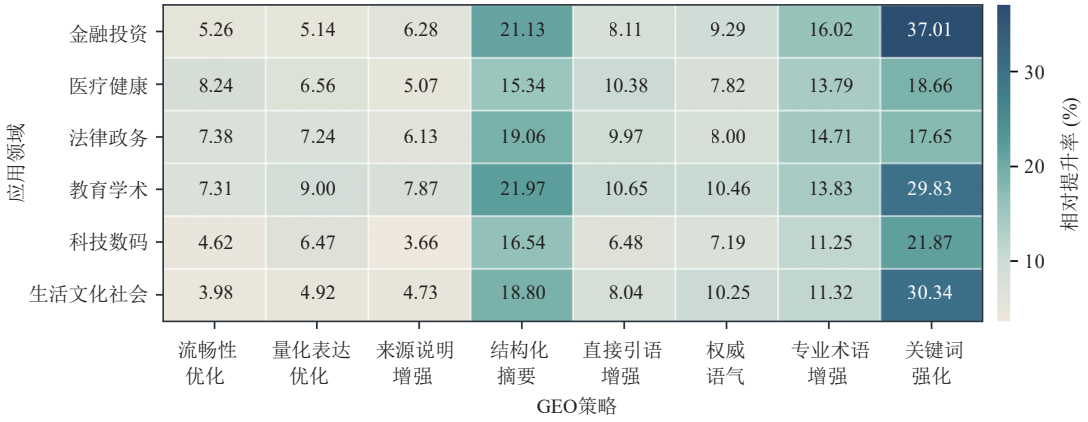


图 11 不同应用领域下 GEO 策略的总体相对提升率

3 种主要策略对领域变化的敏感程度并不相同. 关键词强化在法律政务领域的相对提升率为 17.65%, 在金融投资领域达到 37.01%, 相差 19.36 个百分点, 表现出较大的领域间波动. 这一结果表明, 关键词强化对领域变化较敏感, 这种领域差异可能与各领域原网页对问题核心词汇的覆盖程度不同有关. 当原网页与问题之间的词汇对应较弱时, 关键词强化可能具有更大的提升空间; 当原文已大量使用与问题一致的规范表达时, 其边际收益相对有限.

相比之下, 结构化摘要在各领域中的相对提升率介于 15.34%–21.97% 之间, 专业术语增强介于 11.25%–16.02% 之间, 波动明显小于关键词强化. 由此可见, 关键词强化更适合作为高收益的针对性策略, 而结构化摘要和专业术语增强更适合作为中文网页的通用优化方法.

对 8 种策略的相对提升率取平均后, 教育学术、金融投资、生活文化社会、法律政务、医疗健康和科技数码领域的平均提升率分别为 13.87%、13.53%、11.55%、11.27%、10.73% 和 9.76%. 从具体策略的结果来看, 关键词强化的领域间波动较大, 而结构化摘要和专业术语增强的变化相对较小.

为进一步检验上述领域差异是否具有统计显著性, 本文采用 Kruskal-Wallis 检验, 结果如表 3 所示. 本文对 8 种策略及总体结果对应的 9 个检验进行 Benjamini-Hochberg (BH) 多重检验校正. 经校正后, 仅关键词强化在 6 个领域之间存在显著差异 ($H = 38.13$, 校正后 $p < 0.001$), 其余 7 种策略的领域间差异均未达到统计显著水平. 8 种策略总体的领域差异在校正后同样未达到显著水平 ($H = 12.60$, 校正后 $p = 0.123$). 结果表明, 在当前样本下, 多数内容改写策略未检出显著的领域间差异, 而关键词强化对应用领域更为敏感.

表 3 不同 GEO 策略的领域差异显著性检验结果

策略	H	原始 p	BH校正后 p	BH校正后是否显著
流畅性优化	10.25	0.0685	0.1677	否
量化表达优化	7.38	0.1941	0.2114	否
来源说明增强	9.43	0.0931	0.1677	否
结构化摘要	7.68	0.1746	0.2114	否
直接引语增强	7.65	0.1764	0.2114	否
权威语气	7.13	0.2114	0.2114	否
专业术语增强	9.62	0.0866	0.1677	否
关键词强化	**38.13**	**<0.001**	**<0.001**	**是**
8种策略总体	12.60	0.0274	0.1233	否

注: 加粗表示BH校正后$p < 0.05$的检验结果

综合来看, 关键词强化虽然具有最高的整体可见度增益, 但其效果更容易受到应用领域的影响. 相比之下, 结构化摘要和专业术语增强在各领域中均保持较高收益, 且未检出显著的领域间差异, 因此具有更稳定的跨领域适用性. 实际应用中, 可在结构化摘要和专业术语增强等通用策略的基础上, 根据具体领域进一步调整关键词强化程度.

4.5 候选来源位置与可优化空间

针对 RQ4, 本节比较不同候选来源位置下的基线可见度和内容改写后的提升幅度. 不同候选来源位置的基线可见度和总体相对提升率如图 12 所示. 来源 1 的基线综合可见度最高, 为 23.03%, 但总体相对提升率最低, 仅为 9.15%; 来源 4 的基线综合可见度最低, 为 14.71%, 总体相对提升率则最高, 达到 14.17%. 这说明候选来源位置会影响网页的初始可见度, 即靠前来源更容易被模型优先采用, 但其可见度已经处于较高水平, 后续改写带来的边际收益较小; 中后位置来源虽然初始表现较弱, 却拥有更大的优化空间.

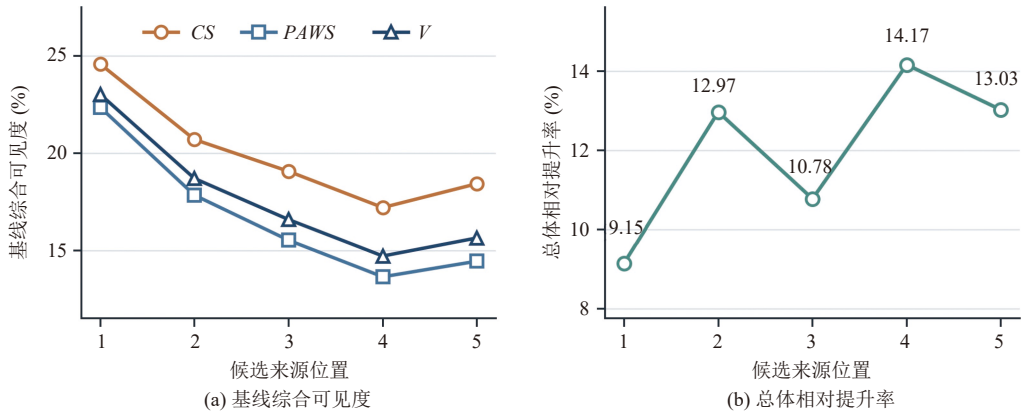


图 12 不同候选来源位置的基线综合可见度与总体相对提升率

来源 2–来源 5 的总体相对提升率分别为 12.97%、10.78%、14.17% 和 13.03%, 并未随位置后移而持续上升. 这表明, 候选位置影响了网页的初始可见度, 但最终提升幅度还取决于网页与问题的相关程度、来源间的信息重合以及关键信息是否便于模型识别和提取. 因此, 候选位置主要带来不同的初始可见度, 网页内容则会进一步影响这种位置优势能否延续或发生改变.

不同策略在 5 个候选来源位置上的相对提升率如图 13 所示. 关键词强化在来源 4 上的提升率达到 31.85%, 明显高于来源 1 的 20.20%; 专业术语增强在来源 1 和来源 5 上的提升率分别为 11.31% 和 15.15%, 来源说明增强则分别为 3.39% 和 7.23%. 多数策略在中后位置上的收益高于来源 1, 说明内容改写能够部分弥补候选顺序带来的劣势, 使更多回答内容来自原本较少被采用的网页.

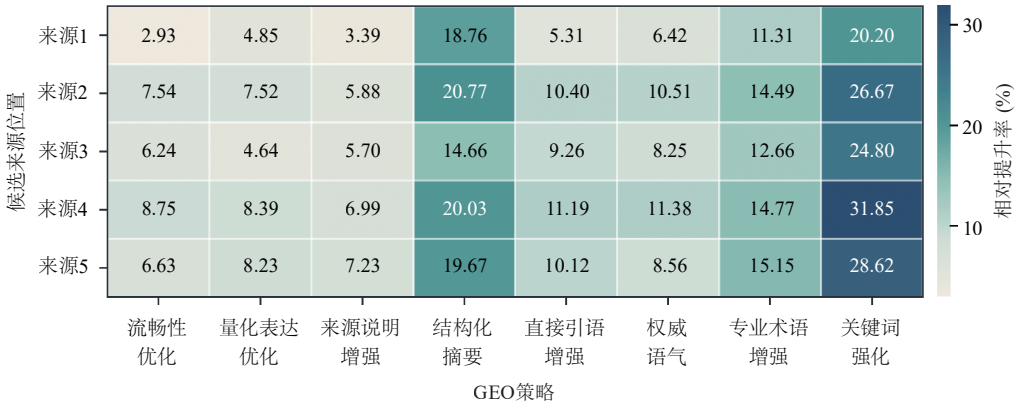


图 13 不同候选来源位置下 GEO 策略的总体相对提升率

关键词强化、结构化摘要和专业术语增强在所有候选位置上均取得正向提升, 进一步说明这些策略的效果并不局限于某一候选位置. 候选顺序会影响模型对不同网页的关注和使用, 但不会决定最终的来源分配. 只要网页与问题之间的对应关系更加明确, 关键信息更加集中, 术语表达更加规范, 即使处于较后位置, 也能够通过内容优化提高来源可见度.

由此可见, 内容优化能够在一定程度上弥补候选位置带来的劣势. 对于已经处于靠前位置的网页, 优化重点应放在巩固现有优势, 避免信息分散削弱其初始可见度; 对于处于中后位置的网页, 则应优先采用关键词强化、结构化摘要和专业术语增强, 通过提高内容的可识别性和可提取性获得更大的可见度提升. 候选位置影响网页的初始可见度, 内容优化则影响其后续提升幅度.

4.6 综合可见度权重的稳健性

主实验取 $\alpha = 0.7$, 即内容贡献率 *PAWS* 和来源引用率 *CS* 的权重分别为 70% 和 30%. 为检验策略排序是否依赖这一权重设置, 本文进一步将 α 调整为 0.5、0.6、0.7、0.8 和 0.9, 并重新计算各策略的综合可见度及相对提升率. 不同权重设置下的结果如图 14 所示.



图 14 不同综合可见度权重下 GEO 策略的总体相对提升率

在 5 种权重设置下, 8 种策略的排序始终保持一致, 关键词强化、结构化摘要和专业术语增强稳定占据前 3 位. 由于这 3 种策略在 *CS* 和 *PAWS* 上均保持领先, 无论综合可见度更侧重来源引用还是内容贡献, 其相对优势都不会改变.

随着 *PAWS* 权重提高, 关键词强化的相对提升率由 25.04% 增至 26.71%, 结构化摘要由 18.14% 增至 19.49%, 专业术语增强也呈现相同趋势. 这表明 3 种策略在提高来源引用的同时, 更明显地提高了目标网页对回答内容的贡献. 随着综合可见度中内容贡献权重的提高, 这 3 种策略的优势更加明显. 可见, *CS* 和 *PAWS* 两项指标均支持主要策略的排序结果, 权重变化仅影响各策略的提升幅度, 不会改变整体结论.

4.7 改写质量与改写模型稳健性分析

4.7.1 人工改写质量评审

在人工改写质量评审中, 两名评审者分别对中性预处理和策略改写结果进行独立评分. 中性预处理主要考察事实一致性、预处理任务完成度和文本质量; 策略改写主要考察事实一致性、策略实现程度和文本质量. 两类任务的人工评价结果如表 4 所示.

人工评审结果表明, 中性预处理和策略改写在事实一致性、任务完成程度和文本质量等方面均获得较高评分, 两名评审者之间的评分差异总体较小. 尤其是策略改写在事实一致性、策略实现程度和文本质量这 3 个维度上均表现稳定, 说明本文的改写过程能够在保持文本质量和事实一致性的同时实现预期策略目标.

表 4 中性预处理与策略改写的人工质量评审结果

任务	评价维度	均分	分差≤1的比例 (%)	平均绝对分差
中性预处理	事实一致性	4.783	100.00	0.200
	预处理任务完成度	4.367	100.00	0.667
	文本质量	4.758	100.00	0.450
策略改写	事实一致性	4.691	100.00	0.165
	策略实现程度	4.682	99.79	0.156
	文本质量	4.942	100.00	0.046

4.7.2 替代改写模型稳健性验证

为进一步检验主要实验结论是否依赖 DeepSeek-v4-pro 及其生成的固定改写版本, 本文在人工质量评审所使用的 60 个问题子集上, 改用 Qwen-3.8-max 作为改写模型, 重新开展内容改写和回答生成实验. 对于 60 个问题对应的 300 篇候选网页, 首先使用 Qwen-3.8-max 重新进行统一中性预处理, 并在此基础上分别生成 8 种策略改写版本; 回答生成阶段仍采用主实验中的豆包、DeepSeek、通义千问和文心一言这 4 个平台, 并保持候选来源、输入顺序、回答提示和生成参数不变. 为避免不同改写模型生成的网页版本在同一实验条件中混合, Qwen-3.8-max 实验的基线条件统一使用其生成的 5 篇中性预处理网页, 策略条件仅将目标来源替换为对应的策略改写版本. 两种改写模型在这 60 个问题子集上的策略效果比较如表 5 所示.

表 5 DeepSeek-v4-pro 与 Qwen-3.8-max 下 GEO 策略效果的稳健性比较

策略	DeepSeek相对提升率 (%)	Qwen相对提升率 (%)	DeepSeek排名	Qwen排名
关键词强化	21.62	23.30	1	1
结构化摘要	14.91	19.13	2	2
专业术语增强	11.92	13.44	3	3
权威语气	9.30	8.96	4	5
直接引语增强	8.91	9.48	5	4
流畅性优化	6.57	6.15	6	7
来源说明增强	4.99	6.59	7	6
量化表达优化	4.81	3.79	8	8

更换改写模型后, 8 种策略均继续取得正向提升, 其中关键词强化、结构化摘要和专业术语增强在两种改写模型下均保持前 3 名, 且三者排序完全一致. 两套策略排序的 Spearman 等级相关系数达到 0.952 4, 仅权威语气与直接引语增强、流畅性优化与来源说明增强分别出现相邻名次交换. 虽然两种改写模型下各策略的具体提升幅度有所不同, 但 8 种策略的提升方向保持一致, 整体排序也较为稳定, 说明更换改写模型并未改变各策略之间的相对效果, 主要实验结论具有较好的稳健性.

5 中文 GEO 与英文 GEO 的差异

第 3 节构建了中文 GEO 评测框架, 第 4 节从整体效果、模型平台、应用领域和候选来源位置等方面分析了不同内容改写策略的表现. 本节进一步回答研究问题 5.

● RQ5 (中英文差异): 中文 GEO 与英文 GEO 的策略效果有何差异?

其中, 第 5.1 节比较中英文 GEO 的实验设定. 第 5.2 节比较相近策略的实验结果. 第 5.3、5.4 节在上述结果基础上进一步讨论中文 GEO 的作用机制和内容优化原则.

5.1 中英文 GEO 的实验对比

针对 RQ5, 本节首先从实验语料、生成系统、基线设置、目标来源选择和策略边界等方面比较中文 GEO 与英文 GEO 的实验设定. 中文 GEO 与英文 GEO 均通过改写网页内容, 考察目标网页在生成式回答中的来源可见度变化. 英文 GEO 主要基于英文问题和 Google 搜索结果, 在基于 GPT-3.5 构建的生成式引擎上开展实验[8], 并使

用 Perplexity.ai 进行验证; 中文 GEO 则面向中文问题和百度搜索结果, 在豆包、DeepSeek、通义千问和文心一言这 4 个国内大模型平台上进行评测. 两者在实验语料、生成系统、基线设置、目标来源选择和策略边界等方面的具体差异如表 6 所示.

表 6　英文 GEO 与中文 GEO 实验设定对比

比较维度	英文GEO	中文GEO
查询与网页	英文问题及Google前5条搜索结果	中文问题及百度前5条搜索结果
主要评测样本	GEO-bench中的1 000条问题	6个领域的600个中文问题
生成系统	GPT-3.5、Perplexity.ai	豆包、DeepSeek、通义千问和文心一言
基线设置	未经改写的网页正文	经统一中性预处理的网页正文
目标来源	每个问题随机选择1篇目标网页	5篇候选网页依次作为目标来源
生成次数	每种设置重复生成5次	每种设置重复生成5次
策略边界	允许补充外部非原文内容	仅调整和突出原文已有信息, 不新增事实
重点分析	领域、网页排名、组合策略	领域、网页排名、模型平台

两者最主要的区别是策略允许修改的内容范围不同. 英文 GEO 中的部分策略可以向网页补充关键词、统计信息、来源和引语[8], 改写后不仅表达形式发生变化, 网页的信息量和证据密度也可能随之增加. 中文 GEO 则将所有改写限制在原文事实范围内, 只对已有信息进行调整、突出和重新组织, 不补充外部事实. 英文 GEO 考察网页经过内容增强后能够获得多大收益, 中文 GEO 关注的则是在事实内容不变的情况下, 表达方式本身如何影响模型对网页的选择和使用.

在基线设置中, 英文 GEO 直接使用未经改写的网页正文作为基线[8], 中文 GEO 则先对所有候选网页进行统一的中性改写, 再将目标网页替换为相应的策略版本. 若直接将策略版本与原始正文比较, 结果中会同时包含语言清洗、格式整理和策略优化带来的影响. 采用中性改写作为基线后, 基线条件和策略条件均经过同一改写模型处理, 两者之间的差异主要来自目标策略所强化的内容特征, 因而更有利于单独考察策略效果.

在目标来源的选择方面, 英文 GEO 为每个问题随机选取一篇网页进行优化[8], 得到不同策略在随机目标来源上的平均效果. 中文 GEO 让 5 篇候选网页依次作为目标来源, 可以分别观察不同候选位置上的基线表现和优化空间. 第 4 节的结果显示, 靠前来源通常具有更高的初始可见度, 中后位置来源则往往能够从内容改写中获得更大的提升. 对所有位置逐一评测, 可以避免随机抽样后的结果掩盖不同候选位置之间的差异.

在生成系统方面, 英文 GEO 主要使用自建生成式引擎, 并通过 Perplexity.ai 验证策略能否迁移到实际系统; 中文 GEO 同时评测 4 个国内大模型平台, 重点分析同一策略在不同平台上的效果与波动. 实验结果表明, 各平台对主要策略的排序较为一致, 但提升幅度存在明显差异. 这意味着中文 GEO 既存在跨平台通用的内容优化规律, 也需要根据具体平台调整策略重点.

综上, 中文 GEO 形成了不同于英文 GEO 的实验范式. 英文 GEO 更强调通过扩充和改写网页内容提高可见度, 中文 GEO 则在统一的事实边界下, 进一步考察词汇表达、信息结构和术语使用等特征各自对来源可见度的影响.

5.2　中英文 GEO 的策略效果差异

为进一步回答 RQ5, 本节比较中文 GEO 与英文 GEO 中相近内容改写策略的效果. 本文按照优化目标对中英文策略进行对应比较, 英文 GEO 采用位置调整词数 (position-adjusted word count, PAWC)[8], 中文 GEO 采用内容贡献率 (*PAWS*). 两项指标均综合考虑归因于目标来源的内容篇幅及其所在位置, 主要区别在于英文 GEO 按照英文词数计算, 中文 GEO 按照中文字符数计算. 为便于比较, 下文将二者统一称为内容贡献率. 英文 GEO 各策略的内容贡献率变化[8]与中文 GEO 的实验结果如表 7 所示.

两组结果呈现出明显不同的策略排序. 英文 GEO 中, 直接引语增强和统计信息添加的提升最为突出, 流畅性优化和来源引用也取得了较高增益, 关键词堆叠则使内容贡献率下降; 中文 GEO 中, 关键词强化取得最高增益, 专业术语增强也取得较高收益, 而直接引语增强、量化表达优化和来源说明增强的提升明显较低. 中文 GEO 新增的

结构化摘要同样取得较高收益, 其内容贡献率提升达到 19.86%, 仅次于关键词强化.

表 7 中英文 GEO 策略实现与内容贡献率变化

中文GEO策略	英文GEO实现	中文GEO实现	英文贡献率 (%)	中文贡献率 (%)
关键词强化	直接增加查询关键词 (关键词堆叠)	强化关键词及相关表达	−8.30	27.17
结构化摘要	无对应策略	结论前置、分层与要点化	—	19.86
专业术语增强	增加主题相关术语 (专业术语)	规范原文已有术语	17.60	14.42
直接引语增强	补充相关引语 (引语添加)	突出原文已有引语	40.90	9.38
权威语气	增强说服力和权威性 (权威语气)	增强专业性保留风险边界	10.40	9.23
量化表达优化	增加统计信息 (统计信息添加)	突出原文已有数字	30.60	6.66
流畅性优化	简化并改善文本表达 (流畅性优化)	优化语序、衔接和重复	28.00	6.21
来源说明增强	补充可信来源 (来源引用)	突出原文已有来源	27.50	5.61

注: 括号中为对应的英文GEO策略名称, 详见第3.2.1节

这种差异首先来自两种研究对策略边界的不同规定. 英文 GEO 中的引语添加、统计信息添加和来源引用策略允许补充网页原本不包含的内容[8], 从而增加可供模型采用的信息和依据, 是这些策略取得较高增益的主要原因. 而中文 GEO 不允许引入外部信息, 这些策略只能利用网页原有材料. 当原文缺少数字、权威来源或可直接引用的表述时, 相应策略便没有足够的操作空间. 因此, 这些策略的效果更依赖原网页是否具备相应的信息基础.

● 直接引语增强. 从数值上看, 在中英文 GEO 中的效果差异较为明显. 英文 GEO 中, 引语添加使内容贡献率相对提升 40.90%, 而中文 GEO 中的相对提升仅为 9.38%. 表面上是中英文 GEO 的区别, 实际上这一差异更可能来自中英文 GEO 的实现方式不同. 英文 GEO 通过补充与主题相关的引语, 为网页增加了新的信息内容和可供模型采用的表述; 中文 GEO 则从原文中抽取语义完整的短句, 突出已有信息, 使其更容易被模型识别和引用. 相比之下, 前者扩充了网页能够提供的信息, 后者更多是在原有内容基础上改善表达形式, 因此两者在内容贡献率上的提升幅度存在明显差异.

● 关键词强化. 在中英文 GEO 中呈现出方向相反的效果. 在英文 GEO 中, 关键词策略使内容贡献率相对下降 8.30%, 而在中文 GEO 中相对提升 27.17%. 这一差异表明, 策略的效果不仅取决于关键词的重复频率, 也与关键词在正文中的组织方式以及语言本身的表达习惯有关. 英文 GEO 采用关键词堆叠, 直接重复问题中的关键词, 容易影响文本的自然性, 因此并未带来来源可见度的提升. 相比之下, 中文 GEO 中的关键词强化并非简单堆叠关键词, 而是在保持正文连贯的基础上, 提高核心关键词及其近义表达的出现频率, 从而加强网页内容与问题之间的词汇对应. 其效果可能与中文的表达习惯有关. 中文并不排斥意义相近的内容重复出现, 例如“语言”“道路”等词本身就由意义相近的成分构成; 在句子和篇章中, 重要概念也常通过重复和前后照应得到强调. 因此, 适度重复可以进一步突出核心信息, 而不会明显破坏表达的自然性. 关键词强化正是利用这一特点, 使核心概念在正文不同位置多次出现, 从而增强模型对网页主题和关键信息的识别. 这可能是关键词策略在中文 GEO 中产生正向效果, 而在英文 GEO 中表现不佳的原因之一.

● 结构化摘要. 结构化摘要的高收益进一步表明, 中英文 GEO 的差异并不只体现在词汇层面. 中文网页常通过较长段落展开背景、条件和解释, 关键结论可能分散在正文不同位置. 结构化摘要通过结论前置、要点化表达和层次划分, 使模型更容易定位和提取回答所需内容. 因此, 中文 GEO 在不增加新事实的情况下, 仅通过重新组织已有信息, 也能够显著提高内容贡献率, 而英文 GEO 更强调增加证据.

由此可见, 中英文 GEO 体现了两条不同的优化路径. 英文 GEO 的主要收益来自增加网页的信息量和证据密度, 中文 GEO 的主要收益则来自加强问题与正文的对应关系, 并降低模型识别和提取关键信息的难度. 因此, 策略效果受到具体改写方式、语言表达特点以及是否允许引入新增信息等因素的共同影响.

5.3 中文 GEO 的作用机制

第 4 节的整体结果、分项指标和多维分析共同指向一个基本规律: 有效的中文 GEO 优化, 本质上是在降低模

型识别、提取和理解网页信息的难度. 关键词强化、结构化摘要和专业术语增强分别在词汇对应、信息组织和术语表达这 3 个方面发挥作用.

● 关键词强化主要提高内容的可识别性. 生成式平台虽然具备语义理解能力, 但在多个主题相近的候选网页之间进行选择时, 问题关键词及其相关表达仍能帮助模型判断网页相关性. 该策略在所有平台和各领域中均位居前列, 这说明词汇对齐不仅影响传统搜索中的相关性匹配, 也会影响生成式模型对候选来源的判断.

● 结构化摘要主要提高内容的可提取性. 该策略对 *PAWS* 的提升高于对 *CS* 的提升, 说明其作用并不只是增加网页被标注为来源的次数, 而是使更多回答内容来自目标网页. 结论前置、小标题和要点化表达减少了关键信息在长文本中的分散程度, 从而提高目标网页在回答内容中的占比.

● 专业术语增强主要提高概念表达的一致性. 将日常表达规范为统一的专业术语, 有助于模型将相关信息对应到明确的领域概念, 减少模型在概念识别和信息整合中的歧义. 该策略的平均增益虽然低于关键词强化和结构化摘要, 但在不同平台和领域之间的波动更小, 说明术语规范是最稳定的中文 GEO 基础策略之一.

5.4 中文 GEO 的内容优化原则

结合中英文比较和第 4 节的实验结果, 中文网页的 GEO 优化可以归纳为以下 4 个方面.

1) 优先增强网页与问题之间的对应关系. 网页标题、核心结论和正文应在保持表达自然的基础上覆盖用户问题中的核心概念, 并合理使用规范名称、常用简称和必要的近义表达.

2) 集中呈现能够直接进入回答的关键信息. 核心结论、主要依据、适用条件和风险提示应放在便于识别的位置, 避免被大量背景内容淹没. 对于篇幅较长的网页, 可以通过结论前置、小标题、要点和分层结构组织内容, 帮助模型快速定位能够直接用于回答的信息.

3) 保持领域概念和专业术语的一致性. 同一概念不宜反复使用含义模糊或不一致的表达, 首次出现时可以同时给出规范名称和常用说法. 对于金融投资、医疗健康和法律政务等领域, 在规范术语的同时, 应保留适用条件、不确定性说明和风险边界, 不能通过增强专业语气将条件性结论改写为确定性判断.

4) 有条件地使用证据类策略. 当原网页已经包含可靠数字、明确出处或具有代表性的直接引语时, 可以通过突出和重组提高这些信息的显著性.

总体而言, 中文 GEO 的核心在于使原有信息与用户问题对应得更直接、组织得更清楚、表达得更稳定. 对于需要同时覆盖多个平台的网页, 可以将专业术语增强和结构化摘要作为基础策略, 再通过关键词强化进一步提高来源可见度.

6 总结和展望

本文围绕中文网页在生成式回答中的来源可见度问题, 构建了面向中文场景的 GEO 评测框架. 该框架覆盖 6 个领域的中文问题及候选网页, 在统一的中性预处理和改写约束下设计 8 种内容改写策略, 并在豆包、DeepSeek、通义千问和文心一言这 4 个国内大语言模型平台上开展对照评测. 本文进一步从来源引用和内容贡献两个维度衡量网页在生成式回答中的表现, 并围绕策略效果、平台差异、领域差异、位置差异和中英文差异这 5 个研究问题展开分析. 实验结果表明, 关键词强化、结构化摘要和专业术语增强整体表现较好, 不同平台、领域和候选来源位置会影响具体提升幅度, 中英文 GEO 在策略效果上也呈现出明显差异.

本文主要研究候选网页确定后的回答生成过程. 未来研究可从 3 个方面展开. 首先, 将评测扩展至查询改写、网页检索、重排序、正文解析和回答生成等完整链路, 分析不同策略在各环节中的作用. 其次, 进一步核验来源标注与回答陈述之间的真实支持关系. 最后, 可扩展策略组合、改写模型和多版本生成等实验条件, 并在语义对齐的中英文问题和网页上开展统一的跨语言对照, 从而更准确地区分语言特点、网页表达习惯和模型平台的影响.

References

[1] Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih WT, Rocktäschel T, Riedel S, Kiela D. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems.

Vancouver: Curran Associates Inc., 2020. 9459–9474.

[2] Kirsten E, Perdekamp JG, Wu QY, Upadhyay M, Gummadi KP, Zafar MB. Characterizing Web search in the age of generative AI. In: Findings of the Association for Computational Linguistics: ACL 2026. San Diego: ACL, 2026. 10827–10848. [doi: 10.18653/v1/2026.findings-acl.526]

[3] Opinions on Deepening the Implementation of the “AI Plus” Initiative. 2025 (in Chinese). http://www.scio.gov.cn/zdgz/jj/202509/t20250901_928364.html

[4] Outline of the 15th Five-year Plan for National Economic and Social Development of the People’s Republic of China. 2026 (in Chinese). https://www.ndrc.gov.cn/fggz/fzzlgh/gjfzgh/202603/U020260317369114704096.pdf

[5] Brin S, Page L. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, 1998, 30(1–7): 107–117. [doi: 10.1016/S0169-7552(98)00110-X]

[6] Berman R, Katona Z. The role of search engine optimization in search marketing. Marketing Science, 2013, 32(4): 644–651. [doi: 10.1287/mksc.2013.0783]

[7] Craswell N, Zoeter O, Taylor M, Ramsey B. An experimental comparison of click position-bias models. In: Proc. of the 2008 Int’l Conf. on Web Search and Data Mining. Palo Alto: ACM, 2008. 87–94. [doi: 10.1145/1341531.1341545]

[8] Aggarwal P, Murahari V, Rajpurohit T, Kalyan A, Narasimhan K, Deshpande A. GEO: Generative engine optimization. In: Proc. of the 30th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. Barcelona: ACM, 2024. 5–16. [doi: 10.1145/3637528.3671900]

[9] Wu YJ, Zhong SS, Kim Y, Xiong CY. What generative search engines like and how to optimize Web content cooperatively. In: Proc. of the 14th Int’l Conf. on Learning Representations. Rio de Janeiro: OpenReview.net, 2026.

[10] Nestaas F, Debenedetti E, Tramèr F. Adversarial search engine optimization for large language models. In: Proc. of the 13th Int’l Conf. on Learning Representations. Singapore: OpenReview.net, 2025.

[11] Xue NW. Chinese word segmentation as character tagging. Int’l Journal of Computational Linguistics & Chinese Language Processing, 2003, 8(1): 29–48.

[12] Zhou LJ, Li BY, Wei ZY, Wong KF. The CUHK discourse TreeBank for Chinese: Annotating explicit discourse connectives for the Chinese TreeBank. In: Proc. of the 9th Int’l Conf. on Language Resources and Evaluation. Reykjavik: ELRA, 2014. 942–949.

[13] Muller B, Wieting J, Clark JH, Kwiatkowski T, Ruder S, Baldini Soares L, Aharoni R, Herzig J, Wang XY. Evaluating and modeling attribution for cross-lingual question answering. In: Proc. of the 2023 Conf. on Empirical Methods in Natural Language Processing. Singapore: ACL, 2023. 144–157. [doi: 10.18653/v1/2023.emnlp-main.10]

[14] Zhang K, He XY, Yao JG. From citation selection to citation absorption: A measurement framework for generative engine optimization across AI search platforms. arXiv:2604.25707, 2026.

[15] Joachims T, Granka L, Pan B, Hembrooke H, Gay G. Accurately interpreting clickthrough data as implicit feedback. In: Proc. of the 28th Annual Int’l ACM SIGIR Conf. on Research and Development in Information Retrieval. Salvador: ACM, 2005. 154–161. [doi: 10.1145/1076034.1076063]

[16] Lewandowski D, Sünkler S, Yagci N. The influence of search engine optimization on Google’s results: A multi-dimensional approach for detecting SEO. In: Proc. of the 13th ACM Web Science Conf. 2021. ACM, 2021. 12–20. [doi: 10.1145/3447535.3462479]

[17] Gyöngyi Z, Garcia-Molina H. Web spam taxonomy. In: Proc. of the 1st Int’l Workshop on Adversarial Information Retrieval on the Web. Chiba: AIRWeb, 2005. 39–47.

[18] Gyöngyi Z, Garcia-Molina H, Pedersen J. Combating Web spam with TrustRank. In: Proc. of the 30th Int’l Conf. on Very Large Data Bases. Toronto: VLDB Endowment, 2004. 576–587.

[19] Ntoulas A, Najork M, Manasse M, Fetterly D. Detecting spam Web pages through content analysis. In: Proc. of the 15th Int’l Conf. on World Wide Web. Edinburgh: ACM, 2006. 83–92. [doi: 10.1145/1135777.1135794]

[20] Bardas N, Mordo T, Kurland O, Tennenholtz M. Automatic document editing for improved ranking. In: Proc. of the 48th Int’l ACM SIGIR Conf. on Research and Development in Information Retrieval. Padua: ACM, 2025. 2779–2783. [doi: 10.1145/3726302.3730168]

[21] Goren G, Kurland O, Tennenholtz M, Raiber F. Ranking-incentivized quality preserving content modification. In: Proc. of the 43rd Int’l ACM SIGIR Conf. on Research and Development in Information Retrieval. ACM, 2020. 259–268. [doi: 10.1145/3397271.3401058]

[22] Raifer N, Raiber F, Tennenholtz M, Kurland O. Information retrieval meets game theory: The ranking competition between documents’ authors. In: Proc. of the 40th Int’l ACM SIGIR Conf. on Research and Development in Information Retrieval. Tokyo: ACM, 2017. 465–474. [doi: 10.1145/3077136.3080785]

[23] Goren G, Kurland O, Tennenholtz M, Raiber F. Ranking robustness under adversarial document manipulations. In: Proc. of the 41st Int’l ACM SIGIR Conf. on Research and Development in Information Retrieval. Ann Arbor: ACM, 2018. 395–404. [doi: 10.1145/3209978.3210012]

[24] Zhang XL, Qu XY, Xie JF, Liu PC, Xiao CG, Wang JZ. AIGC: A survey of AI-generated content technologies. Big Data Research, 2026, 12(1): 146–173 (in Chinese with English abstract). [doi: 10.11959/j.issn.2096-0271.2025071]

[25] Karpukhin V, Oguz B, Min S, Lewis P, Wu L, Edunov S, Chen DQ, Yih WT. Dense passage retrieval for open-domain question answering. In: Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing. ACL, 2020. 6769–6781. [doi: 10.18653/v1/2020.emnlp-main.550]

[26] Izacard G, Grave E. Leveraging passage retrieval with generative models for open domain question answering. In: Proc. of the 16th Conf. of the European Chapter of the Association for Computational Linguistics: Main Volume. ACL, 2021. 874–880. [doi: 10.18653/v1/2021.eacl-main.74]

[27] Shuster K, Poff S, Chen M, Kiela D, Weston J. Retrieval augmentation reduces hallucination in conversation. In: Findings of the Association for Computational Linguistics: EMNLP 2021. Punta Cana: ACL, 2021. 3784–3803. [doi: 10.18653/v1/2021.findings-emnlp.320]

[28] Bohnet B, Tran VQ, Verga P, Aharoni R, Andor D, Baldini Soares L, Ciaramita M, Eisenstein J, Ganchev K, Herzig J, Hui K, Kwiatkowski T, Ma J, Ni J, Sestorain Saralegui L, Schuster T, Cohen WW, Collins M, Das D, Metzler D, Petrov S, Webster K. Attributed question answering: Evaluation and modeling for attributed large language models. arXiv:2212.08037, 2022.

[29] Gao TY, Yen H, Yu JT, Chen DQ. Enabling large language models to generate text with citations. In: Proc. of the 2023 Conf. on Empirical Methods in Natural Language Processing. Singapore: ACL, 2023. 6465–6488. [doi: 10.18653/v1/2023.emnlp-main.398]

[30] Nakano R, Hilton J, Balaji S, Wu J, Ouyang L, Kim C, Hesse C, Jain S, Kosaraju V, Saunders W, Jiang X, Cobbe K, Eloundou T, Krueger G, Button K, Knight M, Chess B, Schulman J. WebGPT: Browser-assisted question-answering with human feedback. arXiv:2112.09332, 2021.

[31] Menick J, Trebacz M, Mikulik V, Aslanides J, Song F, Chadwick M, Glaese M, Young S, Campbell-Gillingham L, Irving G, McAleese N. Teaching language models to support answers with verified quotes. arXiv:2203.11147, 2022.

[32] Liu NF, Zhang TY, Liang P. Evaluating verifiability in generative search engines. In: Findings of the Association for Computational Linguistics: EMNLP 2023. Singapore: ACL, 2023. 7001–7025. [doi: 10.18653/v1/2023.findings-emnlp.467]

[33] Rashkin H, Nikolaev V, Lamm M, Aroyo L, Collins M, Das D, Petrov S, Tomar GS, Turc I, Reitter D. Measuring attribution in natural language generation models. Computational Linguistics, 2023, 49(4): 777–840. [doi: 10.1162/coli_a_00486]

[34] Yue X, Wang BS, Chen ZR, Zhang K, Su Y, Sun H. Automatic evaluation of attribution by large language models. In: Findings of the Association for Computational Linguistics: EMNLP 2023. Singapore: ACL, 2023. 4615–4635. [doi: 10.18653/v1/2023.findings-emnlp.307]

[35] Ding YF, Facciani M, Joyce E, Poudel A, Bhattacharya S, Veeramani B, Aguinaga S, Weninger T. Citations and trust in LLM generated responses. In: Proc. of the 39th AAAI Conf. on Artificial Intelligence. Philadelphia: AAAI, 2025. 23787–23795. [doi: 10.1609/aaai.v39i22.34550]

[36] Huang CY, Wu ZQ, Hu YS, Wang WY. Training language models to generate text with citations via fine-grained rewards. In: Proc. of the 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok: ACL, 2024. 2926–2949. [doi: 10.18653/v1/2024.acl-long.161]

[37] Huang L, Feng XC, Ma WT, Gu YX, Zhong WH, Feng XC, Yu WJ, Peng WH, Tang DY, Tu DD, Qin B. Learning fine-grained grounded citations for attributed large language models. In: Findings of the Association for Computational Linguistics: ACL 2024. Bangkok: ACL, 2024. 14095–14113. [doi: 10.18653/v1/2024.findings-acl.838]

[38] Yu JW, Yang MF, Ding YP, Sato H. Structural feature engineering for generative engine optimization: How content structure shapes citation behavior. arXiv:2603.29979, 2026.

[39] Chen QY, Chen JH, Huang HS, Shao Q, Chen JT, Hua RJ, Xu HX, Wu RJ, Chuan R, Wu J. CC-GSEO-Bench: A content-centric benchmark for measuring source influence in generative search engines. arXiv:2509.05607, 2025.

[40] Wu BN, Mao FY, Lin J, Yang C, Lu JX, Guo YF, Zhang SY, Wu YF, Huang Y, Li F. From experience to skill: Multi-agent generative engine optimization via reusable strategy learning. In: Findings of the Association for Computational Linguistics: ACL 2026. San Diego: ACL, 2026. 43305–43315. [doi: 10.18653/v1/2026.findings-acl.2149]

[41] Yuan JQ, Wang JL, Wang ZH, Sun QY, Wang RJ, Li JX. AgenticGEO: A self-evolving agentic system for generative engine optimization. arXiv:2603.20213, 2026.

[42] Liu ZK, Xu PL. Think before writing: Feature-level multi-objective optimization for generative citation visibility. In: Proc. of the 64th Annual Meeting of the Association for Computational Linguistics. San Diego: ACL, 2026. 20290–20303. [doi: 10.18653/v1/2026.acl-long.929]

[43] Chen T, Guo JW, Li YX, Chen BM, Ren HX, Zhang ZW, Zhang YX, Xia HY, Liang K, Fan ZR. Mind Reader: Latent user demand-

guided content optimization for generative search engine. In: Proc. of the 64th Annual Meeting of the Association for Computational Linguistics. San Diego: ACL, 2026. 40832–40848. [doi: 10.18653/v1/2026.acl-long.1894]

[44] Zhou HY, Chen JJ, Chen XL, Bao J, Chen Z, Liao Y. IF-GEO: Conflict-aware instruction fusion for multi-query generative engine optimization. In: Findings of the Association for Computational Linguistics: ACL 2026. San Diego: ACL, 2026. 27576–27590. [doi: 10.18653/v1/2026.findings-acl.1373]

[45] Puerto H, Gubri M, Green T, Oh SJ, Yun S. C-SEO bench: Does conversational SEO work? In: Proc. of the 39th Int'l Conf. on Advances in Neural Information Processing Systems. San Diego: NeurIPS, 2025.

[46] Vishwakarma R, Kumar S, Jamidar R. What gets cited: Competitive GEO in AI answer engines. In: Proc. of the 49th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Melbourne: ACM, 2026. 4950–4954. [doi: 10.1145/3805712.3808445]

[47] Kim S, Jeong W, Kim S, Lee S, Lee D. SAGEO Arena: A realistic environment for evaluating search-augmented generative engine optimization. In: Proc. of the 32nd ACM SIGKDD Conf. on Knowledge Discovery and Data Mining V. 2. Jeju: ACM, 2026. 2342–2353. [doi: 10.1145/3770855.3818146]

[48] Tian ZH, Chen YH, Tang Y, Liu J, Jia RX. Diagnosing and repairing citation failures in generative engine optimization. arXiv: 2603.09296, 2026.

[49] Emerson T. The second international Chinese word segmentation bakeoff. In: Proc. of the 4th SIGHAN Workshop on Chinese Language Processing. ACL, 2005. 123–133.

[50] Liu NF, Lin K, Hewitt J, Paranjape A, Bevilacqua M, Petroni F, Liang P. Lost in the middle: How language models use long contexts. Trans. of the Association for Computational Linguistics, 2024, 12: 157–173. [doi: 10.1162/tacl_a_00638]

[51] Volcano Engine. Volcano ark model list. 2026 (in Chinese). https://docs.volcengine.com/docs/82379/1330310?lang=zh

[52] DeepSeek. DeepSeek API docs: Your first API call. https://api-docs.deepseek.com/

[53] Baidu AI Cloud. Baidu Qianfan large model service and agent development platform: Model list. 2026 (in Chinese). https://cloud.baidu.com/doc/qianfan/s/rmh4stp0j

[54] Alibaba Cloud. Model studio: Text generation. 2026 (in Chinese). https://help.aliyun.com/zh/model-studio/text-generation-model

[55] Efron B. Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 1979, 7(1): 1–26.

附中文参考文献

[3] 国务院关于深入实施"人工智能+"行动的意见. 2025. http://www.scio.gov.cn/zdgz/jj/202509/t20250901_928364.html

[4] 中华人民共和国国民经济和社会发展第十五个五年规划纲要. 2026. https://www.ndrc.gov.cn/fggz/fzzlgh/gjfzgh/202603/U020260317369114704096.pdf

[24] 张旭龙, 瞿晓阳, 谢骏飞, 刘鹏程, 肖春光, 王健宗. 人工智能生成内容技术综述. 大数据, 2026, 12(1): 146–173. [doi: 10.11959/j.issn.2096-0271.2025071]

[51] 火山引擎. 火山方舟模型列表. https://docs.volcengine.com/docs/82379/1330310?lang=zh

[53] 百度智能云. 百度千帆大模型服务及 Agent 开发平台: 模型列表. https://cloud.baidu.com/doc/qianfan/s/rmh4stp0j

[54] 阿里云. 大模型服务平台百炼: 文本生成. https://help.aliyun.com/zh/model-studio/text-generation-model

作者简介

李彤, 博士, 教授, CCF 高级会员, 主要研究领域为互联网体系结构, 分布式大数据系统, 生成式引擎优化.

吴荣榜, 硕士生, 主要研究领域为大语言模型, 生成式引擎优化.

李清昊, 硕士, 主要研究领域为金融行业营销传播, 用户行为数据分析.

贾贵, 硕士, 主要研究领域为资产管理, 财富管理行业数字化建设, AI 智能体应用.

高翔宇, 博士生, 主要研究领域为智能网络流量分析, 可编程数据平面.

范举, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为智能数据库技术, 人在回路数据准备, 大数据管理与分析.

徐恪, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为新一代互联网体系结构, 网络空间安全, 区块链系统.